# INFERENCE FOR THE INTRINSIC SEPARATION AMONG DISTRIBUTIONS WHICH MAY DIFFER IN LOCATION AND SCALE

By

YAN LING

B.S., Nanjing University of Aeronautics and Astronautics, 1994
M.S., Kansas State University, 2005

AN ABSTRACT OF A DISSERTATION

Submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Statistics

College of Art and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2009

# Abstract

The null hypothesis of equal distributions, $H_0 : F_1 = F_2 = ... = F_K$, is commonly used to compare two or more treatments based on data consisting of independent random samples. Using this approach, evidence of a difference among the treatments may be reported even though from a practical standpoint their effects are indistinguishable, a long-standing problem in hypothesis testing. The concept of *effect size* is widely used in the social sciences to deal with this issue by computing a unit-free estimate of the magnitude of the departure from $H_0$ in terms of a change in location. I extend this approach by replacing $H_0$ with hypotheses $\{H_0^*\}$ that state that the distributions $\{F_i\}$ are possibly different in location and or scale, but *close*, so that rejection provides evidence that at least one treatment has an important practical effect. Assessing statistical significance under $H_0^*$ is difficult and typically requires inference in the presence of nuisance parameters. I will use frequentist, Bayesian and Fiducial modes of inference to obtain approximate tests and carry out simulation studies of their behavior in terms of size and power. In some cases a bootstrap will be employed. I will focus on tests based on independent random samples arising from $K \geq 3$ normal distributions not required to have the same variances to generalize the $K = 2$ sample parameter $\mathrm{P}(X_1 > X_2) = \int F_2(y)F_1(dy)$ and non-centrality type parameters that arise in testing for the equality of means.

# INFERENCE FOR THE INTRINSIC SEPARATION AMONG DISTRIBUTIONS WHICH MAY DIFFER IN LOCATION AND SCALE

By

YAN LING

B.S., Nanjing University of Aeronautics and Astronautics, 1994
M.S., Kansas State University, 2005

A DISSERTATION

Submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Statistics

College of Art and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2009

Approved by:

Major Professor
Paul I. Nelson

# Abstract

The null hypothesis of equal distributions, $H_0 : F_1 = F_2 = ... = F_K$, is commonly used to compare two or more treatments based on data consisting of independent random samples. Using this approach, evidence of a difference among the treatments may be reported even though from a practical standpoint their effects are indistinguishable, a long-standing problem in hypothesis testing. The concept of *effect size* is widely used in the social sciences to deal with this issue by computing a unit-free estimate of the magnitude of the departure from $H_0$ in terms of a change in location. I extend this approach by replacing $H_0$ with hypotheses $\{H_0^*\}$ that state that the distributions $\{F_i\}$ are possibly different in location and or scale, but *close*, so that rejection provides evidence that at least one treatment has an important practical effect. Assessing statistical significance under $H_0^*$ is difficult and typically requires inference in the presence of nuisance parameters. I will use frequentist, Bayesian and Fiducial modes of inference to obtain approximate tests and carry out simulation studies of their behavior in terms of size and power. In some cases a bootstrap will be employed. I will focus on tests based on independent random samples arising from $K \geq 3$ normal distributions not required to have the same variances to generalize the $K = 2$ sample parameter $P(X_1 > X_2) = \int F_2(y)F_1(dy)$ and non-centrality type parameters that arise in testing for the equality of means.

# Table of Contents

# List of Figures

# List of Tables

# Acknowledgements

I would like to express my deepest appreciation to my major professor, Prof. Paul I. Nelson, for suggesting this interesting topic and for his expert guidance in my work. I would also like to thank him for his patience and understanding throughout my pursuit of this degree. He has always cared for his students and has encouraged me to explore the field of statistics more deeply. His dedication and diligence in his work also make a strong impression on my memory.

I also would like to thank Prof. James W. Neill, Prof. Weixing Song, Prof. Robert Burckel (Department of Mathematics), and Prof. Daniel A. Andresen (Department of Computing and Information Sciences) for their valuable suggestions and their willingness to serve on my committee.

Finally, I wish to extend my gratitude to my parents and my family for their everlasting love, encouragement and material and spiritual support. Without their support, I could not have completed my study.

# Chapter 1 Introduction

The commonly used practice of comparing the locations of two or more distributions as a method for assessing how different one treatment is from another can be misleading, even when they have the same shape. Suppose, for example that $X_1$ and $X_2$ are independent, normally distributed random variables, $X_i \sim N(\mu_i, \sigma_i^2)$, with cumulative distribution functions denoted by $\Phi((x-\mu_i)/\sigma_i)$, $i = 1,2$, where $\Phi$ is the standard normal distribution function. Then, if the means are equal

$$\pi_{12} \equiv P(X_1 > X_2) = \Phi((\mu_1 - \mu_2)/\sqrt{(\sigma_1^2 + \sigma_2^2)}) = 1/2\,, \tag{1.1}$$

and can be made arbitrarily close to ½ by adjusting the variances, equal or not, no matter how different the means are. This behavior is illustrated in Table 1 and Figure 1 below. If, for example, $X_1$ and $X_2$ represent yields of two different varieties of wheat, all other things being equal, a farmer would prefer variety one to variety two if $\pi_{12} > 0.95$, regardless of how close the means are. Correspondingly, if $\mu_1$ is a lot bigger than $\mu_2$ but $\pi_{12}$ is close to ½ , the same farmer might prefer variety two to variety one if it is cheaper to plant, grow and bring to market. The failure of $\Delta_{12} \equiv \mu_1 - \mu_2$ to adequately represent the difference between two distributions arises here because $\Delta_{12}$ ignores the variation in the distributions and is not scale invariant. Furthermore, standard tests for the equality of means based on independent random samples are consistent and will declare the distributions to be different when sample sizes are large, no matter how close the means are, as long as they are not identical. Broadly framed, this long-standing issue concerns distinguishing statistical from practical significance. These problems can be somewhat remedied by using what is called an *effect size*, denoted ES, given here by $ES_1 = (\mu_1 - \mu_2)/\sqrt{(\sigma_1^2 + \sigma_2^2)}$ or by $ES_2 = (\mu_1 - \mu_2)/(\sigma_1 + \sigma_2)$, to assess the separation between two normal distributions. Note that $\pi_{12}$ is a monotone increasing function of $ES_1$ and that both decrease rapidly in Table 1 as $\sigma^2 = \sigma_1^2 = \sigma_2^2$ increases. These location-scale invariant effect sizes are examples of what I call *intrinsic separation parameters*.

Table 1. The Relationship Between $\sigma^2$ and $\pi_{12}$ under Normality for Fixed $\mu_1 - \mu_2$

| $\mu_1 - \mu_2$ | $\sigma^2$ | $ES_1$ | $ES_2$ | $\pi_{12}$ | Figure |
|---|---|---|---|---|---|
| 1 | 0.1 | 2.23607 | 1.58114 | 0.98733 | Figure1.1 |
| 1 | 1 | 0.70711 | 0.50000 | 0.76025 | Figure1.2 |
| 1 | 10 | 0.22361 | 0.15811 | 0.58847 | Figure1.3 |
| 1 | 100 | 0.07071 | 0.05000 | 0.52819 | Figure1.4 |
| 1 | 1000 | 0.02236 | 0.01581 | 0.50892 | |

Figure1. Separation Between Two Normal Distributions Having Common Variance

$(\mu_1 - \mu_2 = 1)$

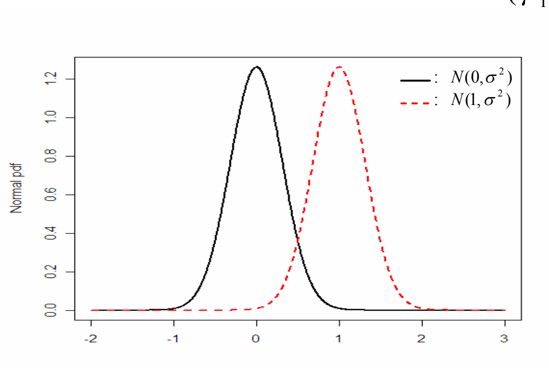



Figure 1.1 $\sigma^2$=0.1

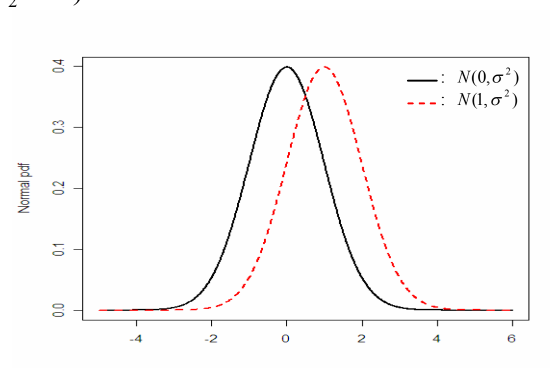



Figure 1.2 $\sigma^2$=1

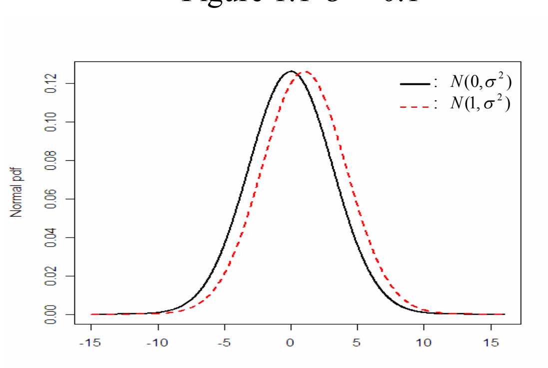



Figure 1.3 $\sigma^2$=10

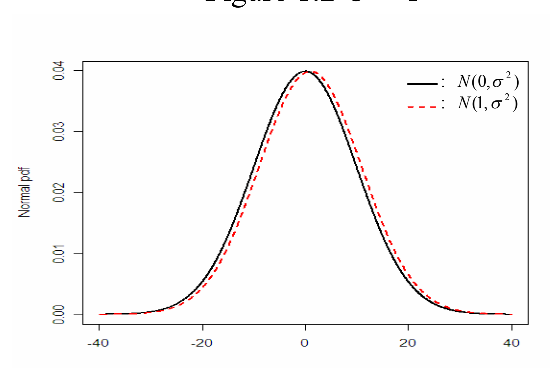



Figure 1.4 $\sigma^2$=100

Definition: Given $K$ continuous distributions, $\underline{F} = \{F_i\}$, a real-valued function IS($\underline{F}$), invariant with respect to location-scale transformations of the form $F_i^{\#}(x) = F_i(ax+b)$, $a > 0$, $i = 1, 2, ..., K$, whose value increases with increasing differences among the $\{F_i\}$ will be called an ***intrinsic separation parameter*** (ISP).

The need for this new terminology arises because the term *effect size* as a measure of separation has been identified almost exclusively and narrowly with differences among locations of distributions of the from $\underline{F} = \{F_i(x) = F((x-\mu_i)/\sigma_i),\ \sigma_i > 0,\ i = 1, 2, ..., K\}$, restricted by the assumption that $\sigma_1 = \sigma_2 = \cdots = \sigma_K$. Specifically, for comparing the means of two normal distributions having the same standard deviation $\sigma$, the effect size is commonly taken to be ES = $(\mu_1 - \mu_2)/\sigma$. The literature in this area contains no work on inference for $ES_1$ and $ES_2$, which are special cases of my research. The usefulness of an ISP depends on the extent to which its values have meaning to the user.

My research develops and explores inference for intrinsic separation among two or more distributions. The parametric part of my work will expand the scope of conclusions that can be reached by comparing distributions of the form $\{F_i(x) = F((x-\mu_i)/\sigma_i),\ \sigma_i > 0,\ i = 1, 2, ..., K\}$, without requiring that they have the same scale, based on a realization of independent random samples $\{\underline{X}_i = \underline{x}_i = (x_{i1}, x_{i2}, ..., x_{in_i})\}$, with sample means and variances, denoted $\{(\bar{X}_i, S_i^2)\}$. Letting IS($\underline{F}$) denote an intrinsic separation parameter of interest, I propose testing

$$H_0\text{: IS}(\underline{F}) \le \Psi \quad \text{vs} \quad H_1\text{: IS}(\underline{F}) > \Psi\ , \tag{1.2}$$

where $\Psi$ is a user input value. The value of $\Psi$ ideally represents the smallest magnitude of separation among the distributions as measured by a particular IS($\underline{F}$) which is of practical importance. An alternative, classical approach to this issue, as presented in Hodges and Lehmann (1954) and Lehmann and Romano (2005), is to use three hypotheses; $H^{(0)}$: IS($\underline{F}$) = 0, $H^{(1)}$: 0 < IS($\underline{F}$) $\le \Psi$, and $H^{(2)}$ : IS($\underline{F}$) $> \Psi$, where $H^{(1)}$

represents an *indifference zone* in the parameter space. Constructing tests for these three hypotheses can be very difficult and I will not pursue this approach. Inverting a test for (1.2) also yields confidence sets for IS($\underline{F}$), which can be used to provide a data-based assessment of the magnitude of the separation among the distributions. In the one-way analysis of variance, $K \geq 2$, based on independent random samples from normal distributions having means $\{\mu_i$, i = 1,2,…, $K\}$ and the same unknown variance $\sigma^2$, IS($\underline{F}$) is commonly, if only implicitly, taken to be the non-centrality parameter of the $F$ - test, given by

$$\mathrm{IS}_{\mathrm{AOV\_EQ}}(\underline{F}) = \sum_{i=1}^{K} n_i(\mu_i - \bar{\mu})^2 / \sigma^2 , \tag{1.3}$$

where $\bar{\mu} = \sum_{i=1}^{K} n_i\mu_i / N$ , $N = \sum_{i=1}^{K} n_i$ . The standard procedure here too is to take $\Psi = 0$ so that neither $H_0$, which is never true in practice, nor $H_1$ accurately indicates just how different the distributions are. In this case, testing (1.2) with $\Psi > 0$ is relatively easy to carry out, as is described for $K = 2$ in Hodges and Lehmann (1954) and for $K \geq 3$ in Murphy and Myors (2004). Specifically, since the F- family has monotone likelihood ratio with respect to its non-centrality parameter, having observed $F = F_{obs}$ , rejecting $H_0$ if $F_{obs}$ exceeds the 1- $\alpha$ quantile of the non-central F-distribution with degrees of freedom $K$-1 and $N$-$K$ and non-centrality parameter $\Psi$ , yields a consistent, size $\alpha$ test.

Now, consider the one-way normal theory analysis of variance described above, where the variances $\{\sigma_i^2\}$ need not be equal. An obvious extension of (1.3) is the intrinsic separation parameter defined by

$$\mathrm{IS}_{\mathrm{AOV}}(\underline{F}) = \sum_{i=1}^{K} [n_i(\mu_i - \bar{\mu})^2 / \sigma_i^2], \tag{1.4}$$

where now $\bar{\mu} = \sum_{i=1}^{K} \left(n_i\mu_i / \sigma_i^2\right) / \sum_{i=1}^{K} \left(n_i / \sigma_i^2\right)$. A reasonable, exact test for (1.4) is not available. However, the random variable

$$\mathrm{Q}(\underline{\sigma}) = \sum_{i=1}^{K} w_i(\bar{X}_i - \bar{X}_w)^2 / \sum_{i=1}^{K} w_i , \tag{1.5}$$

where $w_i = n_i / \sigma_i^2$ and $\overline{X}_w = \sum_{i=1}^{K} w_i \overline{X}_i / \sum_{i=1}^{K} w_i$, has a non-central chi-square distribution with $K$-1 degrees of freedom and non-centrality parameter equal to $IS_{AOV}(\underline{F})$ as given in (1.4). For $K$ = 2, using $IS_{AOV}(\underline{F})$ and taking $\Psi$ = 0 reduces to the Behrens - Fisher problem. Rice and Gains (1989) developed a Fiducial test based on $Q(\underline{\sigma})$ and Krishnamoorthy, Lu, and Mathew (2007) carried out a parametric bootstrap test based on $Q(\underline{S}^2) = (S_1^2, S_2^2, \ldots, S_K^2)$, both for the equality of means, $\Psi$ = 0. I will extend both of these results and develop Bayesian and Fiducial approaches to cover inference for $\Psi$ > 0 in future research .

One area I will focus on concerns samples from normal distributions and the class of intrinsic separation parameters given by

$$IS_{LIN}(\underline{F}) = \sum_{i=1}^{K} l_i \mu_i / \sqrt{\sum_{i=1}^{K} l_i^2 \sigma_i^2}\ , \tag{1.6}$$

where $l = (l_1, l_2, \ldots, l_K)$ is a user input vector of constants that sum to 0. Note that for independent random samples from normal distributions, $\Phi$ [$IS_{LIN}(\underline{F})$] = $P(\sum_{i=1}^{K} l_i X_i > 0)$ is one way of extending $\pi_{ij} \equiv P(X_i > X_j), i \neq j$, defined above, from $K$ = 2 to $K$ = 3 or more distributions. Two other ways to accomplish this without assuming a particular form for the underlying distributions are by using the ISP's given by

$$IS_{MAX}(\underline{F}) = Max\{\pi_{ij}\}, \tag{1.7}$$

or

$$IS_{AV}(\underline{F}) = 2\sum_{i<j} Max\{\pi_{ij}, \pi_{ji}\}/(K(K-1)). \tag{1.8}$$

Absolutely no treatment effect occurs for (1.7) and (1.8) when $\Psi$ = ½ and when $\Psi$ = 0 for (1.6). Taking $\Psi = \pi >$ ½ for (1.7) and (1.8) and $\Psi = \Phi^{-1}(\pi)$ for (1.6) denotes increasing separation as $\pi$ approaches 1. The separation in (1.6) may be viewed as one-sided, as dictated by the choice of the coefficients $\{l_i\}$, and two-sided for (1.7) and (1.8). Under normality, no matter what the variances may be, when all the means are equal, $IS_{AOV}$(F) and $IS_{LIN}$(F) are 0, the minimum value of $IS_{AOV}$(F). In calibrating $IS_{LIN}$(F), I assume that the coefficients $\{l_i\}$ are chosen so that increasing $IS_{LIN}(F)$ corresponds to increasing separation. Thus, for example, $\{l_i\} = \{-1/2, -1/2, 1\}$ indicates a preference for showing that there is high probability that $X_3$ is greater than $(X_1+X_2)/2$. But, if all the means are equal, (1.7)-(1.8) are guaranteed to equal to ½ , their minimum value, only if the distributions are also symmetric. These ISP's can be used to assess treatment effects that involve both means and variances for skewed as well as symmetric distributions. ISP's (1.7) and (1.8) are also invariant with respect to an increasing transformation of the data. I will describe below results I have obtained on inference for (1.6), (1.7) and (1.8).

Perng, et al. (1989) and Kemp, et al. (1993) constructed normal theory and nonparametric tests, respectively, for IS($\underline{F}$) = $\xi_p^{(1)} - \xi_{1-p}^{(2)}$ and $\Psi > 0$, when $K = 2$, where $\xi_p^{(i)}$ is the $p^{\text{th}}$ quantile of $F_i$, $0 < p < 1/2$. Note that under normality, IS($\underline{F}$) = $ES_2$ with $\Psi = \Phi^{-1}(1-p)$. Here, I will attempt to extend this approach to three or more distributions.

Effect size (ES) is a key concept that has been thoroughly discussed by Cohen (1988) in his book *Statistical Power Analysis for the Behavioral Sciences (2nd ed)*. In general, effect size means "the degree to which the null hypothesis is false." It measures the degree of departure from the null hypothesis. Cohen notes that the powers of many commonly used tests are functions of sample size ($n$), the population effect size (ES), and the desired size $\alpha$. It is possible in principle to solve for any of the four values (power, $n$, ES, $\alpha$) given the other three. Cohen calibrates the practical importance of an estimated effect size as being small, medium or large by relating it to power. I will develop a similar calibration for ISP's and develop a scheme for estimating the sample sizes necessary for my tests to have desired power at specified alternatives in future research.

All of the problems I will study involve inference in the presence of nuisance parameters. Pivotals, when they exist, provide direct solutions to this difficult problem. Specifically, suppose that interest lies in inference for $\Gamma(\underline{\theta}) = \Gamma$. A quantity of the form $Q = Q(\Gamma, Data)$ is pivotal if its distribution $P_Q$ is free of $\underline{\theta}$ when $\Gamma(\underline{\theta}) = \Gamma$. A test of $H_0 : \Gamma = \Gamma_0$ can then be carried out at type I error rate $\alpha$ by rejecting $H_0$ if

$Q_0 = Q(\Gamma_0, Data) \in C$ , where $P_{H_0}(Q_0 \in C) = \alpha$ . Such pivotals do not exist for my problems. Instead, in Chapter 2, I will base tests on functions of the parameters and data of the form $Q = Q(\Gamma(\underline{\theta}), \underline{\Upsilon}(\underline{\theta}), Data)$, whose distributions under $H_0 : \Gamma = \Gamma_0$, denoted $F_Q(\cdot \mid \eta(\underline{\theta}))$ ,are known up to the additional nuisance parameter $\eta(\underline{\theta})$. Let $\hat{\Upsilon}$ and $\hat{\eta}$ be estimates of $\Upsilon$ and $\eta$, respectively, obtained from $Data$. Then, given $Q(\Gamma_0, \hat{\Upsilon}, Data) = q_{obs}$, I will construct and investigate an approximate p-value defined as 1-$F_Q(q_{obs} \mid \hat{\eta})$. This approach is related to what Bayarri and Berger (2000) termed *the plug-in method*. I will also use a parametric bootstrap to calibrate the distribution of a likelihood ratio test. This will require solving the difficult problem of finding estimates of $\underline{\Upsilon}$ constrained by $H_0$.

In Chapter 3, with $F_Q(\cdot \mid \eta(\underline{\theta})) = F_Q(\cdot)$, I will compute p-values = p( $\underline{\Upsilon}$ ) as functions of the nuisance parameters $\underline{\Upsilon}$ and "average" these over Fiducial and Bayes Posterior distributions on $\underline{\Upsilon}$. Berger and Selke(1987) assert that the posterior probability of $H_0$ is a better measure of the evidence in the data than a p-value. Accordingly, I will also use Fiducial and Bayes Posterior distributions to evaluate the probability of the null hypothesis given the data and investigate the use of these values as evidence for choosing between $H_0$ and $H_1$. Although some of the problems I present in Chapter 4 fall outside of the scope of inference for ISP's, they are interesting related issues I worked on while preparing this dissertation. In Chapter 5, I present a simulation study based on the level and power for comparing those p-values in testing $IS_{LIN}(\underline{F})$. At the same time, a simulation study based on the level and power for testing $IS_{MAX}(\underline{F})$ and $IS_{AV}(\underline{F})$ are also presented in Chapter 5. A summary and conclusions are presented in Chapter 6.

# Chapter 2 Frequentist Tests for $IS_{LIN}(\underline{F})$ Assuming Normality

## 2.1 A Plug-In Test

To review some of the ideas presented in the introduction, suppose we have $K$ normal distributions, $\{X_i \sim N(\mu_i, \sigma_i^2), i = 1,2,...,K\}$, and independent random samples $\{ x_{ij}, j = 1,2,\ldots, n_i \}$ from each, $N = \sum n_i$. For $K = 2$, as noted above,

$$P(X_1 > X_2) = \Phi((\mu_1 - \mu_2)/\sqrt{(\sigma_1^2 + \sigma_2^2)}) \equiv \pi_{12}, \tag{2.1}$$

is a location-scale invariant measure of the extent to which the distribution of $X_1$ lies above the distribution of $X_2$. Specifically, if $\pi_{12}$ is close to 1.0, *most* independent copies of $X_1$ will be larger than *most* independent copies of $X_2$. Consider tests of the form

$$H_0 : \pi_{12} = \pi \text{ vs } H_1 : \pi_{12} > \pi, \tag{2.2}$$

where $\pi$ is a proportion at least 0.5. If $\pi = 0.5$ and $\sigma_1 = \sigma_2$, the pooled t-test provides an exact size $\alpha$ test of (2.2). If $\pi = 0.5$ and $\sigma_1 \neq \sigma_2$, (2.2) is the familiar Behrens-Fisher problem, for which there is no reasonable, exact size $\alpha$ test. In this case, Welch's test (Welch 1938) is an approximate size $\alpha$ test. The Mann-Whitney test is only guaranteed to have its nominal size for (2.2) when $F_1 = F_2$ under $H_0$ and hence $\pi = 0.5$. My goals here are to extend the concept of *separation* given in (2.1) to the case of $K \geq 3$ normal distributions and develop tests for the corresponding generalizations of (2.2) that do not require an assumption of equal variances. I will call a test having approximate p-value '*p*', *a nominal size $\alpha$ test* if the null hypothesis is rejected whenever $p \leq \alpha$. Simulation can then be used to check if the actual type I error rate is close to its nominal value.

To generalize (2.1) so as to define separation among $K \geq 3$ distributions, let $\underline{X}^T = (X_1, X_2, ..., X_K)$, $\underline{l}^T = (l_1, l_2, ..., l_K) \neq \underline{0}$, $\sum_{i=1}^{K} l_i = 0$, which is needed for location invariance, $\frac{1}{2} \leq \pi < 1$, where $\underline{X}^T$ denotes the transpose of $\underline{X}$ and the components of the vector $\underline{l}$ are user input constants. Separation can then be defined by a preference for hypothesis $H_1$ over hypothesis $H_0$, where,

$$
\begin{aligned}
H_0 &: \ P(\underline{l}^T \underline{X} > 0) \leq \pi \,, \\
H_1 &: \ P(\underline{l}^T \underline{X} > 0) > \pi \,. 
\end{aligned}
\tag{2.3}
$$

Increasing separation corresponds to increasing $\pi$. Since $\underline{l}^T \underline{X}$ is distributed $N(\sum_{i=1}^{K} l_i \mu_i, \ \sum_{i=1}^{K} l_i^2 \sigma_i^2)$, letting $\Phi(n_\pi) = \pi$, (2.3) can be expressed as:

$$
H_0: \ \Delta(\underline{\mu}, \underline{\sigma}) \ \leq \ n_\pi \,, \quad H_1: \ \Delta(\underline{\mu}, \underline{\sigma}) \ > \ n_\pi \,, \tag{2.4}
$$

where $\Delta(\underline{\mu}, \underline{\sigma}) \equiv \sum l_i \mu_i / \sqrt{\sum l_i^2 \sigma_i^2} = IS_{LIN}(\underline{\mathrm{F}})$.

Constructing tests for (2.4) requires dealing with the nuisance parameters $\rho_i = \sigma_i^2 / \sigma_1^2$, $i = 2,3,...,K$. We begin by noting that since $\sum l_i \bar{X}_i \sim N(\sum l_i \mu_i, \ \sum l_i^2 \sigma_i^2 / n_i)$, we have that

$$
Z \equiv \frac{\sum l_i \bar{X}_i - \sum l_i \mu_i}{\sqrt{\sum l_i^2 \sigma_i^2 / n_i}} \sim N(0,1),
$$

$$
W \equiv \sum \{(n_i - 1) S_i^2 / \sigma_i^2\} \sim \chi^2_{(N-K)},
$$

and $W$ and $Z$ are independent. Hence,

$$T = \frac{\sum l_i \overline{X}_i / \sqrt{\sum l_i^2 \sigma_i^2 / n_i}}{\sqrt{\frac{\sum (n_i - 1) S_i^2 / \sigma_i^2}{(N-K)}}} \tag{2.5}$$

$$= \frac{Z + \delta}{W} \sim t'(N-K, \delta(\underline{\mu}, \underline{\sigma}, \underline{n}) \equiv \sum l_i \mu_i / \sqrt{\sum l_i^2 \sigma_i^2 / n_i}),$$

where $t'(\nu, \delta)$ denotes a non-central t-distribution with non-centrality parameter $\delta$ and $\nu$ degrees of freedom. Dividing numerator and denominator of $T$ by $\sigma_1^2$, setting $\rho_1 = 1$ and letting $\underline{\rho} = (1, \rho_2, ..., \rho_K)$, $T$ can be expressed as

$$T = T(\underline{\rho}) = \frac{\sum l_i \overline{X}_i / \sqrt{\sum l_i^2 \rho_i \sigma_1^2 / n_i}}{\sqrt{\frac{\sum (n_i - 1) S_i^2 / \rho_i \sigma_1^2}{(N-K)}}} = \frac{\sum l_i \overline{X}_i}{\sqrt{\frac{(\sum l_i^2 \rho_i / n_i)(\sum (n_i - 1) S_i^2 / \rho_i)}{(N-K)}}} \ . \tag{2.6}$$

Let $t_{obs}(\underline{\rho})$ be the observed value of $T(\underline{\rho})$ and $N = \sum_{i=1}^{K} n_i$ . If $\{ \rho_i \}$ were known, a p-value for $H_0 : \delta(\underline{\mu}, \underline{\sigma}, \underline{n}) \le \delta_0$, $H_1 : \delta(\underline{\mu}, \underline{\sigma}, \underline{n}) > \delta_0$ could be defined by

$$p(\underline{\rho}) = p - value(\underline{\rho}) = P(\ T \ge t_{obs}(\underline{\rho})\ ), \tag{2.7}$$

where $T \sim t'(df = N-K,\ \delta = \delta_0)$. Now, letting $n_{(1)} = \min\{n_i\}$ and $n_{(K)} = \max\{n_i\}$, we have that under $H_0$ given in (2.4),

$$\sqrt{n_{(1)}} \Delta(\mu, \sigma) \le \delta(\mu, \sigma, n) \le \sqrt{n_{(K)}} \Delta(\mu, \sigma) \ .$$

Since a non-central t-distribution has monotone likelihood ratio in its non-centrality parameter, computing $p(\rho)$ in (2.7) by taking $T \sim t'(N-K, \sqrt{n_{(K)}}(n_\pi))$ provides a conservative test of (2.4), which is exact size $\alpha$ if $n_{(1)} = ... = n_{(K)} = n$. To handle the realistic case where $\underline{\rho}$ is not known, estimate its components by $\hat{\rho}_i = S_i^2 / S_1^2$ , $i = 2,3,\ldots,K$ and plug $\underline{\hat{\rho}}^T = (1, \hat{\rho}_2, \ldots, \hat{\rho}_K)$ into (2.6), yielding a test statistic

$$T(\underline{\hat{\rho}}) = \frac{\sum l_i \overline{X}_i}{\sqrt{\frac{(\sum l_i^2 \hat{\rho}_i / n_i)(\sum (n_i - 1) S_i^2 / \hat{\rho}_i)}{(N-K)}}} = \frac{\sum l_i \overline{X}_i}{\sqrt{\sum l_i^2 S_i^2 / n_i}} \quad . \tag{2.8}$$

An approximate p-value can then be defined as

$$p(\underline{\hat{\rho}}) = \text{p} - \text{value}(\underline{\hat{\rho}}) = P(T \geq t_{obs}(\underline{\hat{\rho}})) \ . \tag{2.9}$$

Where $T \sim t'(df', \sqrt{n_{(K)}}(n_\pi))$ and $df'$ is given by the approximation due to Satterthwaite(1946),

$$df' = \frac{(\sum_{i=1}^{K} \frac{l_i^2}{n_i} s_i^2)^2}{\sum_{i=1}^{K} \left[ (\frac{l_i^2}{n_i} s_i^2)^2 / (n_i - 1) \right]} \quad . \tag{2.10}$$

Based on a preliminary simulation (given in the Appendix A, Figure A.1), the distributions of p-values given in (2.9) appear, as desired, to be approximately uniform under the null hypothesis if the sample sizes are equal. But, for unequal sample sizes, especially for cases where the range of the sample sizes is large, these p-values have a highly skewed distribution under $H_0$.

This situation can be improved by using an estimate of the non-centrality parameter $\delta$ instead of an upper bound. To carry this out, first, rewrite the non-centrality parameter $\delta$ as follows

$$\delta = \sum l_i \mu_i / \sqrt{\sum l_i^2 \sigma_i^2 / n_i}$$

$$= \frac{\sum l_i \mu_i}{\sqrt{\sum l_i^2 \sigma_i^2}} \frac{\sqrt{\sum l_i^2 \sigma_i^2}}{\sqrt{\sum l_i^2 \sigma_i^2 / n_i}} \tag{2.11}$$

$$= \Delta \frac{\sqrt{\sum l_i^2 \sigma_i^2}}{\sqrt{\sum l_i^2 \sigma_i^2 / n_i}},$$

so that under $H_0$, $\delta \le n_\pi \frac{\sqrt{\sum l_i^2 \sigma_i^2}}{\sqrt{\sum l_i^2 \sigma_i^2 / n_i}}$. Again using the fact that the noncentral t-distribution has monotone likelihood ratio in its noncentrality parameter, p-values may be computed using $\delta = n_\pi \frac{\sqrt{\sum l_i^2 \sigma_i^2}}{\sqrt{\sum l_i^2 \sigma_i^2 / n_i}}$, assuming that the population variances were known. Accordingly, again letting $\{ S_i^2 \}$ denote the sample variances, define an estimate of $\delta$ by

$$\hat{\delta} = n_\pi \frac{\sqrt{\sum l_i^2 S_i^2}}{\sqrt{\sum l_i^2 S_i^2 / n_i}}.$$

Using $\nu = df'$ (the Satterthwaite approximation, (2.10) ), results in an estimated p-value, given by:

$$p-value(\underline{S}^2) \equiv P(\ T \ge t_{obs}(\underline{S}^2)\ )\ , \tag{2.12}$$

where $T \sim t'(df',\ \hat{\delta})$.

Preliminary simulations (given in the Appendix A, Figures A.2-A.3) indicate that this approach appears to yield p-values that are approximately uniformly distributed under $H_0$ and tests that have good power, except when the range of sample sizes is very large. A full simulation study investigating the size and power of this test is conducted in Chapter 5. I will investigate its robustness with respect to departures from normality and the presence of outliers in future research.

## 2.2 A Likelihood Ratio Test

The log-likelihood function for $K$ independent random normal distributions is given by

$$\log L(\underline{\mu},\underline{\sigma}^2) = -\sum_{i=1}^{K}\frac{n_i}{2}\log(2\pi\sigma_i^2) - \sum_{i=1}^{K}\sum_{j=1}^{n_i}(x_{ij}-\mu_i)^2/2\sigma_i^2 . \tag{2.13}$$

This log-likelihood is maximized by $\log(L(\underline{\hat{\mu}},\underline{\hat{\sigma}}^2)$, where $\underline{\hat{\mu}}$ and $\underline{\hat{\sigma}}^2$ are the well known maximum likelihood estimators. Carrying out a likelihood ratio test requires maximizing the log-likelihood constrained by $H_0$ in (2.3), which can be difficult to carry out since this constrained likelihood is a very complicated function in a typically high dimensional space. For example, Buot, et al. (2007) and Drton (2008) show that for the Behrens-Fisher problem, $K = 2$ and $\pi = 0.50$, the likelihood function can have multiple modes under $H_0$. As a first step in addressing this problem, I develop a Jacobi type algorithm for finding a local maximum which uses a Lagrange multiplier to incorporate the boundary constraint

$$\frac{\sum_{i=1}^{K} l_i\mu_i}{\sqrt{\sum_{i=1}^{K} l_i^2\sigma_i^2}} = n_\pi . \tag{2.14}$$

Form the function D ( $\underline{\mu}$ $\underline{\sigma}^2, \lambda$ ) defined by

$$\begin{aligned} \mathrm{D} &= \log L(\underline{\mu},\underline{\sigma}^2) + \lambda\cdot\Big(\sum_{i=1}^{K} l_i\mu_i \Big/ \sqrt{\sum_{i=1}^{K} l_i^2\sigma_i^2} - n_\pi\Big) \qquad (2.15) \\ &= -\sum_{i=1}^{K}\frac{n_i}{2}\log(2\pi\sigma_i^2) - \sum_{i=1}^{K}\sum_{j=1}^{n_i}(x_{ij}-\mu_i)^2/2\sigma_i^2 + \lambda\cdot\Big(\sum_{i=1}^{K} l_i\mu_i \Big/ \sqrt{\sum_{i=1}^{K} l_i^2\sigma_i^2} - n_\pi\Big), \end{aligned}$$

where $\lambda$ is constant and is the Lagrange multiplier of the constraint (2.14). Taking partial derivatives and setting them equal to zero, yields the equations

$$\frac{dD}{d\mu_i} = 0 \quad \text{for } i = 1,2,\ldots K,$$

$$\frac{dD}{d\sigma_i^2} = 0 \quad \text{for } i = 1,2,\ldots K,$$

Some simplification yields the following likelihood equations,

$$\frac{\lambda l_i}{\sqrt{\sum_{j=1}^{K} l_j^2 \hat{\sigma}_j^2}} - \frac{n_i \overline{x}_i - n_i \hat{\mu}_i}{\hat{\sigma}_i^2} = 0 \quad , \quad i = 1,2,\ldots K, \tag{2.16}$$

$$\frac{n_i}{2\hat{\sigma}_i^2} - \frac{\sum_{j=1}^{n_i} (x_{ij} - \hat{\mu}_i)^2}{2(\hat{\sigma}_i^2)^2} - \frac{l_i^2 \lambda \sum_{j=1}^{K} l_j \hat{\mu}_j}{2(\sum_{j=1}^{K} l_j^2 \hat{\sigma}_j^2)^{3/2}} = 0 \quad , \; i = 1,2,\ldots K. \tag{2.17}$$

Equation (2.16) becomes

$$\hat{\mu}_i = \overline{x}_i - \frac{\lambda l_i \hat{\sigma}_i^2}{n_i \sqrt{\sum_{j=1}^{K} l_j^2 \hat{\sigma}_j^2}} \quad . \tag{2.18}$$

Now, multiply (2.18) by $l_i$ on both sides and sum. Then,

$$\sum_{i=1}^{K} l_i \hat{\mu}_i = \sum_{i=1}^{K} l_i \overline{x}_i - \lambda \sum_{i=1}^{K} \frac{l_i^2 \hat{\sigma}_i^2}{n_i \sqrt{\sum_{j=1}^{K} l_j^2 \hat{\sigma}_j^2}},$$

$$\Rightarrow \qquad \lambda = \frac{\sqrt{\sum_{i=1}^{K} l_i^2 \hat{\sigma}_i^2}}{\sum_{i=1}^{K} \frac{l_i^2 \hat{\sigma}_i^2}{n_i}} \left( \sum_{i=1}^{K} l_i \overline{x}_i - \sum_{i=1}^{K} l_i \hat{\mu}_i \right) , \qquad \text{by (2.14)}$$

$$\Rightarrow \qquad \lambda = \frac{\sqrt{\sum_{i=1}^{K} l_i^2 \hat{\sigma}_i^2}}{\sum_{i=1}^{K} \frac{l_i^2 \hat{\sigma}_i^2}{n_i}} \left( \sum_{i=1}^{K} l_i \overline{x}_i - n_\pi \sqrt{\sum_{i=1}^{K} l_i^2 \hat{\sigma}_i^2} \right).$$

Plug the above equation into (2.18), which then can be written as

$$\hat{\mu}_i = \bar{x}_i + \frac{l_i \hat{\sigma}_i^2 [n_\pi \sqrt{\sum_{i=1}^{K} l_i^2 \hat{\sigma}_i^2} - \sum_{i=1}^{K} l_i \bar{x}_i]}{n_i \sum_{i=1}^{K} \frac{l_i^2 \hat{\sigma}_i^2}{n_i}} \tag{2.19}$$

Next, from equation (2.14), equation (2.17) becomes:

$$\frac{n_i}{\hat{\sigma}_i^2} - \frac{\sum_{j=1}^{n_i} (x_{ij} - \hat{\mu}_i)^2}{(\hat{\sigma}_i^2)^2} - \frac{\lambda l_i^2 n_\pi^3}{(\sum_{i=1}^{K} l_i \mu_i)^2} = 0,$$

$$\Rightarrow \qquad n_i \hat{\sigma}_i^2 - \sum_{j=1}^{n_i} (x_{ij} - \hat{\mu}_i)^2 - \frac{\lambda l_i^2 n_\pi^3 (\hat{\sigma}_i^2)^2}{(\sum_{i=1}^{K} l_i \mu_i)^2} = 0.$$

From equation (2.16), the above equation can be rewritten as:

$$n_i \hat{\sigma}_i^2 - \sum_{j=1}^{n_i} (x_{ij} - \hat{\mu}_i)^2 - \frac{l_i n_\pi^3 \hat{\sigma}_i^2}{(\sum_{i=1}^{K} l_i \mu_i)^2} (n_i \bar{x}_i - n_i \hat{\mu}) \sqrt{\sum_{i=1}^{K} l_i^2 \hat{\sigma}_i^2} = 0.$$

$$\Rightarrow \qquad n_i \hat{\sigma}_i^2 - \sum_{j=1}^{n_i} (x_{ij} - \hat{\mu}_i)^2 - \frac{l_i n_\pi^2 \hat{\sigma}_i^2}{\sum_{i=1}^{K} l_i \mu_i} (n_i \bar{x}_i - n_i \hat{\mu}) = 0.$$

$$\Rightarrow \qquad \hat{\sigma}_i^2 = \frac{\sum_{j=1}^{n_i} (x_{ij} - \hat{\mu}_i)^2}{n_i - \frac{l_i n_\pi^2 (n_i \bar{x}_i - n_i \hat{\mu}_i)}{\sum_{i=1}^{K} l_i \hat{\mu}_i}} \quad , \quad i = 1,2,\ldots K. \tag{2.20}$$

Therefore, under the boundary constraint (2.14), candidate maximum likelihood estimators of $\{(\mu_i, \sigma_i^2)\}$ are given by the solution of equations (2.19) and (2.20). In Section 2.5 below I present an iterative algorithm for solving these equations. Henceforth, following the usual practice, I will refer to these solutions as maximum likelihood estimators, denoted MLE's.

There are $2K$ unknown parameters and solving this system is difficult. A likelihood ratio test can be carried out by rejecting $H_0$ if

$$\Lambda = \frac{\sup_{H_0}(L(\underline{\mu},\underline{\sigma}^2)}{L(\underline{\hat{\mu}},\underline{\hat{\sigma}}^2)} \le c \quad , \tag{2.21}$$

where $c$ is calibrated so that the test has nominal size $\alpha$. The performance of this test and comparisons to other tests I derive are investigated by simulation in Chapter 5. Letting $\lambda = -2\ln\Lambda$, from the results of the simulation study described later (Appendix B, Result 4), the asymptotic distribution of $\lambda$ under $H_0$ appears in this nonstandard case not to have a chi-square distribution with one degree of freedom. Drton (2009) studies a variety of nonstandard cases and notes that this behavior can result from the nature of the local geometry of the parameter space. It's not even clear whether $\Lambda$ has a limiting distribution and if so whether that distribution is free of the true parameter. This issue needs further study.

Jaber and Cox, in an unpublished manuscript, derived a likelihood ratio test for the two-sided Behrens-Fisher problem. Specifically, for $K = 2$, let $l = (1,-1)$, and $n_\pi = 0$; the hypotheses (2.3) becomes $H_0 : \mu_1 = \mu_2$ against the alternative $H_1 : \mu_1 \ne \mu_2$. Under this null hypothesis, using $\mu$ to denote the common unspecified mean under $H_0$, the maximum likelihood estimators of $\mu, \sigma_1^2, \sigma_2^2$ are the solutions to likelihood equations, given implicitly by

$$\hat{\mu} = \frac{n_1\bar{x}_1\hat{\sigma}_2^2 + n_2\bar{x}_2\hat{\sigma}_1^2}{n_1\hat{\sigma}_2^2 + n_2\hat{\sigma}_1^2}$$

$$\hat{\sigma}_1^2 = S'^2_1 + (\bar{x}_1 - \hat{\mu})^2$$

$$\hat{\sigma}_2^2 = S'^2_2 + (\bar{x}_2 - \hat{\mu})^2$$

where $S'^2_1 = \dfrac{\sum_{j=1}^{n_1}(x_{1j} - \bar{x}_1)^2}{n_1}$ and $S'^2_2 = \dfrac{\sum_{j=1}^{n_2}(x_{2j} - \bar{x}_2)^2}{n_2}$ .

Hence, the LRT statistic is given by:

$$\Lambda = \left( \frac{\hat{\sigma}_1^2}{S'^2_1} \right)^{n_1/2} \left( \frac{\hat{\sigma}_2^2}{S'^2_2} \right)^{n_2/2} \quad . \tag{2.22}$$

Letting $\lambda = -2\ln\Lambda$, the asymptotic distribution of $\lambda$ under $H_0$ is a chi-square distribution with one degree of freedom. Jaber and Cox argued that the size of the LRT using $\hat{\sigma}_1^2$, $\hat{\sigma}_2^2$ (MLE), denoted by ($LRT_1$), is less than that of using $S_1^2, S_1^2$ (sample variances), denoted by ($LRT_2$), and based on a simulation study both tests have sizes close to the nominal significance level . However, $LRT_2$ has slightly higher power than $LRT_1$, and in most cases, $LRT_2$ compares favorably with regard to size and power to the Welch -Aspin test (Welch (1947) and Aspin (1948)), which has size extremely close to the nominal significance level. Furthermore, asymptotic results show that there is some relationship between the generalized likelihood ratio test and the most commonly used test statistic, denoted by $V$ below, for the Behrens-Fisher problem:

$$V = \frac{\overline{X}_1 - \overline{X}_2}{(S_1^2/n_1 + S_2^2/n_2)^{1/2}} . \tag{2.23}$$

## 2.3 Simulation Example

Suppose, for example, we have three independent, normally distributed random variables $\{X_i\}_{i=1}^3$ with means $\mu_i$ ($i$ =1,2,3) , standard deviations $\sigma_i$ ($i$ =1,2,3), and independent random sample of sizes $\{n_i\}$, respectively, and that we want to carry out the following test:

$$H_0 : P(X_3 > \frac{(X_1 + X_2)}{2}) \le \pi = 0.9$$
$$H_a : P(X_3 > \frac{(X_1 + X_2)}{2}) > \pi = 0.9$$

### 2.3.1 Satterthwaite Approximation (Conservative) Test

Test statistic: $T = \dfrac{\sum l_i \overline{X}_i}{\sqrt{\sum l_i^2 S_i^2 / n_i}}$ where $l_i = (-1/2,\ -1/2,\ 1)$

P-value = P($T \geq T_{obs}$), where $T \sim$ t' (df', $\sqrt{n(k)}\ n_\pi$), $\Phi(n_\pi) = \pi$ . Using the Satterthwaite approximation given in (2.10), we have that:

$$df' = \frac{\left(\frac{l_1^2}{n_1} s_1^2 + \frac{l_2^2}{n_2} s_2^2 + \frac{l_3^2}{n_3} s_3^2\right)^2}{\frac{\left(\frac{l_1^2}{n_1} s_1^2\right)^2}{(n_1 - 1)} + \frac{\left(\frac{l_2^2}{n_2} s_2^2\right)^2}{(n_2 - 1)} + \frac{\left(\frac{l_3^2}{n_3} s_3^2\right)^2}{(n_3 - 1)}}.$$

Simulation results based on 10000 iterations are summarized in Figure A.1 of Appendix A, where it can be seen that the p-values are approximately uniformly distributed under $H_0$, except when sample sizes are far apart.

### 2.3.2 Satterthwaite Approximation (Estimate) Test

Test statistic: $T = \dfrac{\sum l_i \overline{X}_i}{\sqrt{\sum l_i^2 S_i^2 / n_i}}$ where $l_i = (-1/2,\ -1/2,\ 1)$

P-value = P($T \geq T_{obs}$), where $T \sim$ t' (df', $\delta = \dfrac{\sqrt{\sum l_i^2 s_i^2}}{\sqrt{\sum l_i^2 s_i^2 / n_i}} n_\pi$**),** $\Phi(n_\pi) = \pi = 0.9$

Simulation studies for this test are given in Appendix A, Figures A.2-A.3, where closer approximations to uniformly distributed p-values under the null hypotheses than in Result 1 are evident and powers approaching 1 are obtained as the ISP increases.

## 2.4 Parametric Bootstrap Tests

If the maximum likelihood estimators under the composite null hypothesis (2.4) could be found, then a parametric bootstrap could be used to calibrate the critical regions for the LRT statistic $\Lambda$ in (2.21), as follows. Stated in general terms, let $G_N(\cdot, P_{\underline{\tau}})$ denote the distribution of a generic test statistic $T = T(\underline{X}_N)$, where the observable $\underline{X}_N$ is generated from a probability law $P_{\underline{\tau}}$, which depends on an unknown parameter $\underline{\tau}$ . Suppose that large values of $T$ favor $H_1$ over $H_0$, statements about $\underline{\tau}$. An approximate size $\alpha$ parametric bootstrap test would reject $H_0$ when an observed value of $T$ exceeds the $1-\alpha$ quantile of $\hat{G}_N(\cdot, P_{\hat{\underline{\tau}}_{0N}})$, where $\hat{\underline{\tau}}_{0,N}$ is an estimate of $\underline{\tau}_0$, a value of $\underline{\tau}$ constrained by $H_0$, and $\hat{G}_N(\cdot, P_{\hat{\tau}_{oN}})$ is an estimate of $G_N(\cdot, P_{\hat{\tau}_{oN}})$ obtained from data generated from $P_{\hat{\tau}_{0N}}$ . Simulation results presented later indicate that this procedure works reasonably well. The following theorem gives conditions under which this parametric bootstrap test will be asymptotically size $\alpha$ in terms of a distance $d$ which metrizes convergence in distribution.

**Theorem 1**. (Lehmann and Romano, 2005) Let $\underline{X}_N$ be generated from a probability law $P \in \mathrm{P}_0$. Assume the following triangular array convergence: $d(P_N, P) \to 0$ and $P \in \mathrm{P}_0$ implies $G_N(.,P_N)$ converges weakly to $G(.,P)$, with $G(.,P)$ continuous. Moreover, assume $\hat{Q}_N$ is an estimator of $P$ based on $\underline{X}_N$ which satisfies $d(\hat{Q}_N, P) \to 0$ in probability whenever $P \in \mathrm{P}_0$. Then, for $P = P_{\underline{\tau}_0} \in P_0$,

$$P\{T_N > G_N^{-1}(1-\alpha, \hat{Q}_N)\} \to \alpha \text{ as } N \to \infty. \tag{2.24}$$

To apply this theorem with $T$ being the likelihood ratio statistic given in (2.21), taking $\hat{Q}_N = \hat{G}_N(T, P_{\hat{\tau}_{oN}})$ is a natural choice. As stated at the end of Section 2.3, although this choice appears to work well in my simulation study, it’s not known if the conditions of

Theorem 1 hold here. However, the existence of maximum likelihood estimators $\{\hat{\underline{\tau}}_{0,N}\}$ such that

$$\{\hat{\underline{\tau}}_{0,N}\} \to \underline{\tau}_0 \text{ in } P_{\underline{\tau}_0} \text{ probability,} \tag{*}$$

is part of a condition that makes $\{P_{\underline{\tau}}\}$ what Drton (2009) calls a *regular statistical model* and provides a heuristic justification for the asymptotic validity of the parametric bootstrap test procedure. I will now establish that (*) holds in a variety of cases by following Silvey(1975) and showing that *a.e.* under $P_{\theta_0}$, for any sufficiently small $\delta$,

$$\lim_N \sup\{\sup_{\|\underline{\tau}-\underline{\tau}_0\|=\delta} (\underline{\tau}-\underline{\tau}_0)^T \dot{l}_N(\underline{\tau})\} < 0\,, \tag{**}$$

where $l_N(\underline{\tau}) = \log(f(x\,|\,\underline{\tau}))$ is the log-likelihood, and $\dot{l}_N(\underline{\tau}) = \dfrac{\partial l_N(\underline{\tau})}{\partial \underline{\tau}}$ is its vector of partial derivatives.

### 2.4.1 Applying (**) to Testing the Mean of a Normal Distribution

<u>**Verification 2.4.1**</u>:

Let $\{X_i\}$ be independent~N($\mu,\theta$), Var($X_i$) = $\theta$ > 0. Beginning with a simple illustration, suppose we want to test:

$$H_0 : \mu = \mu_0,\ H_a : \mu > \mu_0$$

Following (**), we have to show here that *a.e.* under $H_0$, for any sufficiently small $\delta$,

$$\lim_N \sup\{\sup_{|\theta-\theta_0|=\delta} (\theta-\theta_0)\dot{l}_N(\theta)\} < 0\,.$$

This will follow from the first order Taylor expansion

$$\dot{l}_N(\theta) = \dot{l}_N(\theta_0) + (\theta-\theta_0)\ddot{l}_N(\bar{\theta})\,,$$

where: $\bar{\theta}$ represents, here and from now on, an appropriate intermediate value,

$$|\theta - \bar{\theta}| \le |\theta - \theta_0| = \delta\,,$$

$$\dot{l}_N(\theta) = -N/2\theta + \sum (X_i - \mu_0)^2 / 2\theta^2\,,$$

$$\ddot{l}_N(\theta) = N/2\theta^2 - \sum (X_i - \mu_0)^2 / \theta^3\,.$$

Then, using the strong law of large numbers, *a.e.* under $H_0$, we have for $0 < \delta < \theta_0$:

$$\begin{aligned}(\theta - \theta_0)\dot{l}_N(\theta) &= N[o(1) + (\theta - \theta_0)^2(\bar{\theta} - 2(\bar{\theta} + o(1)))/2\bar{\theta}^3] \\ &= N[o(1) + \delta^2(-\bar{\theta} + o(1))/2\bar{\theta}^3] \\ &\le N[o(1) + \delta^2(-\theta_0 + \delta + o(1))/2(\theta_0 + \delta)^3] \\ &\to -\infty, \text{ as } N \to \infty\end{aligned}$$

which completes the verification.

### 2.4.2 Applying (**) to the Behrens - Fisher Problem

**Verification 2.4.2**:

Suppose there are two independent random variables $\{X_i\}_{i=1}^{n_1} \sim \mathrm{N}(\mu_1, \theta_1)$, Var($X_i$) = $\theta_1 > 0$, and $\{Y_j\}_{j=1}^{n_2} \sim \mathrm{N}(\mu_2, \theta_2)$, Var($Y_i$) = $\theta_2 > 0$ and we want to test

$$H_0 : \mu_1 = \mu_2\,, \quad H_a : \mu_1 \ne \mu_2\,.$$

To show that under $H_0$, $\hat{\underline{\theta}} \to \underline{\theta}_0$ , in probability, Theorem 1 can again be invoked by showing that *a.e.* under $H_0$, for any sufficiently small $\delta$ , $N = n_1 + n_2$ and $n_1 / N \to \lambda \in (0,1)$,

$$\lim_N \sup\{ \sup_{\|\underline{\theta} - \underline{\theta}_0\| = \delta} (\underline{\theta} - \underline{\theta}_0)^T \dot{l}_N(\underline{\theta})\} < 0\,,$$

where $\underline{\theta}^T = (\mu, \theta_1, \theta_2)$, $\underline{\theta}_0^T = (\mu_0, \theta_{10}, \theta_{20})$. In this case, we have

$$\dot{l}_N(\underline{\theta}) = \left( \frac{\partial l_N(\underline{\theta})}{\partial \mu}, \quad \frac{\partial l_N(\underline{\theta})}{\partial \theta_1}, \quad \frac{\partial l_N(\underline{\theta})}{\partial \theta_2} \right)^T$$

$$\frac{\partial l_N(\underline{\theta})}{\partial \mu} = \frac{\sum_{i=1}^{n_1}(X_i-\mu)}{\theta_1} + \frac{\sum_{j=1}^{n_2}(Y_j-\mu)}{\theta_2}, \qquad \frac{\partial^2 l_N(\underline{\theta})}{(\partial \mu)^2} = -n_1/\theta_1 - n_2/\theta_2,$$

$$\frac{\partial l_N(\underline{\theta})}{\partial \theta_1} = -\frac{n_1}{2\theta_1} + \frac{\sum_{i=1}^{n_1}(X_i-\mu)^2}{2\theta_1^2}, \qquad \frac{\partial^2 l_N(\underline{\theta})}{(\partial \theta_1)^2} = \frac{n_1}{2\theta_1^2} - \frac{\sum_{i=1}^{n_1}(X_i-\mu)^2}{\theta_1^3},$$

$$\frac{\partial l_N(\underline{\theta})}{\partial \theta_2} = -\frac{n_2}{2\theta_2} + \frac{\sum_{j=1}^{n_2}(Y_j-\mu)^2}{2\theta_2^2}, \qquad \frac{\partial^2 l_N(\underline{\theta})}{(\partial \theta_2)^2} = \frac{n_2}{2\theta_2^2} - \frac{\sum_{j=1}^{n_2}(Y_j-\mu)^2}{\theta_2^3},$$

$$\frac{\partial^2 l_N(\underline{\theta})}{\partial \mu \partial \theta_1} = -\frac{\sum_{i=1}^{n_1}(X_i-\mu)}{\theta_1^2}, \qquad \frac{\partial^2 l_N(\underline{\theta})}{\partial \mu \partial \theta_2} = -\frac{\sum_{j=1}^{n_2}(Y_j-\mu)}{\theta_2^2}, \qquad \frac{\partial^2 l_N(\underline{\theta})}{\partial \theta_1 \partial \theta_2} = 0$$

I prove the consistency of the MLE's under $H_0$ for the Behrens - Fisher problem based on the following principles.

(1) Show that a.e each of the diagonal terms of $\ddot{l}_N(\bar{\theta}) \to -\infty$ for all sufficiently small $\delta > 0$.

(2) Show that the off-diagonal terms are smaller than the diagonal terms in absolute value as $\delta \to 0$.

***Notes***: For sufficiently small $\delta > 0$,

(i) For {x,y,z} such that $x^2+y^2+z^2 = \delta^2$, positive values $\{a, b, c\}$, then:

$ax^2 + by^2 + cz^2 \geq \delta^2 Min\{a,b,c\}$

(ii) $\theta_{i0}/2 \leq \bar{\theta}_i \leq 3\theta_{i0}/2$, $i = 1,2$

$\mu_0/2 \leq \bar{\mu} \leq 3\mu_0/2$

(1) Clearly, $\frac{\partial^2 l_N(\bar{\theta})}{(\partial\mu)^2} = -n_1/\bar{\theta}_1 - n_2/\bar{\theta}_2$ satisfies (1) and using the strong law of large numbers,

$$\frac{\partial^2 l_N(\bar{\theta})}{(\partial\theta_1)^2} = \frac{n_1}{2\bar{\theta}_1^2} - \frac{\sum_{i=1}^{n_1}(X_i-\bar{\mu})^2}{\bar{\theta}_1^3} \quad = \quad \frac{n_1}{\bar{\theta}_1^2}[1/2 - \frac{\sum_{i=1}^{n_1}(X_i-\bar{\mu})^2/n_1}{\bar{\theta}_1}]$$

$$= \quad \frac{n_1}{\bar{\theta}_1^2}[1/2 - \frac{\sum_{i=1}^{n_1}(X_i-\mu_0)^2/n_1 + (\mu_0-\bar{\mu})^2 + 2(\mu_0-\bar{\mu})(\bar{X}-\mu_0)]}{\bar{\theta}_1}]$$

$$= \quad \frac{n_1}{\bar{\theta}_1^2}[1/2 - \frac{\theta_{10} + (\mu_0-\bar{\mu})^2 + \circ(1)}{\bar{\theta}_1}] = \frac{n_1}{\bar{\theta}_1^2}[\frac{\bar{\theta}_1 - 2\theta_{10} - 2(\mu_0-\bar{\mu})^2 + o(1)}{2\bar{\theta}_1}]$$

$$\le \frac{4n_1}{27\theta_{10}^3}[-1.5\theta_{10} + 2\delta^2 + o(1)]$$

$\to -\infty$, as $n_1 \to \infty$, and as $\delta \to 0$.

Likewise, $\frac{\partial^2 l_N(\bar{\theta})}{(\partial\theta_2)^2} = \frac{n_2}{2\bar{\theta}_2^2} - \frac{\sum_{j=1}^{n_2}(Y_j-\bar{\mu})^2}{\bar{\theta}_2^3} \le \frac{4n_2}{27\theta_{20}^3}[-1.5\theta_{20} + 2\delta^2 + o(1)]$.

$\to -\infty$, as $n_2 \to \infty$, and as $\delta \to 0$.

(2) To verify (2), we have that *a.e.*,

$$\left|\frac{\partial^2 l_N(\bar{\theta})}{\partial\mu\partial\theta_1}/n_1\right| = \frac{\left|\sum_{j=1}^{n_1}(X_j-\bar{\mu})/n_1\right|}{\bar{\theta}_1^2}$$

$$= \left|o(1) + (\mu_0-\bar{\mu})\right|/\bar{\theta}_1^2 \le [o(1)+\delta]/\bar{\theta}_1^2$$

$$\le 2[o(1)+\delta]/\theta_{10}^2.$$

Likewise $\left|\frac{\partial^2 l_N(\bar{\theta})}{\partial\mu\partial\theta_2}/n_2\right| < 2[o(1)+\delta]/\theta_{20}^2$.

Since:

$$(\underline{\theta}-\underline{\theta}_0)^T \dot{l}_N(\underline{\theta}) = (\underline{\theta}-\underline{\theta}_0)^T \dot{l}_N(\underline{\theta}_0) + (\underline{\theta}-\underline{\theta}_0)^T \ddot{l}_N(\underline{\bar{\theta}})(\underline{\theta}-\underline{\theta}_0)$$

$$= (\mu-\mu_0)\frac{\partial l_N(\underline{\theta}_0)}{\partial \mu} + (\theta_1-\theta_{10})\frac{\partial l_N(\underline{\theta}_0)}{\partial \theta_1} + (\theta_2-\theta_{20})\frac{\partial l_N(\underline{\theta}_0)}{\partial \theta_2}$$

$$+ (\mu-\mu_0)^2\frac{\partial^2 l_N(\underline{\bar{\theta}})}{(\partial \mu)^2} + (\theta_1-\theta_{10})^2\frac{\partial^2 l_N(\underline{\bar{\theta}})}{(\partial \theta_1)^2} + (\theta_2-\theta_{20})^2\frac{\partial^2 l_N(\underline{\bar{\theta}})}{(\partial \theta_2)^2}$$

$$+ 2(\mu-\mu_0)(\theta_1-\theta_{10})\frac{\partial^2 l_N(\underline{\bar{\theta}})}{\partial \mu \partial \theta_1} + 2(\mu-\mu_0)(\theta_2-\theta_{20})\frac{\partial^2 l_N(\underline{\bar{\theta}})}{\partial \mu \partial \theta_2} + 2(\theta_1-\theta_{10})(\theta_2-\theta_{20})\frac{\partial^2 l_N(\underline{\bar{\theta}})}{\partial \theta_1 \partial \theta_2},$$

$$(\underline{\theta}-\underline{\theta}_0)^T \dot{l}_N(\underline{\theta}) \le N\Big[-\delta^2 \min\{2(\lambda/\theta_{10}+(1-\lambda)/\theta_{20}), \frac{4\lambda}{27\theta_{10}^3}[1.5\theta_{10}-2\delta^2+o(1)],$$

$$\frac{4(1-\lambda)}{27\theta_{20}^3}[1.5\theta_{20}-2\delta^2+o(1)]\} + o_N(1)+o_\delta(1)\Big]$$

$\to -\infty$, as $N \to \infty$, and as $\delta \to 0$.

where *a.e.* $o_N(1) \to 0$ as $N \to \infty$; $o_\delta(1) \to 0$ as $\delta \to 0$, which completes the verification.

### 2.4.3 Applying (**) to $IS_{LIN}$ (F) under $H_0$ in (2.4)

**<u>Verification 2.4.3</u>**:

Let $\bar{X}_i \sim N(\mu_i, \sigma_i^2/n)$ , $i = 1,2,\ldots, K$, independent. Then the un-restricted parameter space is given by: $\Theta = \{\underline{\theta}^T = (\mu_1, \mu_2, ..., \mu_K, \sigma_1^2, \sigma_2^2, ..., \sigma_K^2)\}$. Here, inference is desired for $\zeta(\theta) = \sum l_i \mu_i / \sqrt{\sum l_i^2 \sigma_i^2} = \zeta$ >0. Therefore, the restricted parameter space is defined by the following:

$$\tilde{\Theta} = \{\underline{\vartheta}^T = (\zeta, \mu_2, ..., \mu_K, \sigma_1^2, \sigma_2^2, ..., \sigma_K^2)\}, \quad K \ge 2.$$

Taking without loss of generality $l_1 \ne 0$,

.

$$\mu_1 = [\zeta\sqrt{\sum l_i^2\sigma_i^2} - \sum_{i=2}^{K} l_i\mu_i]/l_1$$

Let $\underline{\vartheta}^T = (\zeta, \mu_2, ..., \mu_K, \theta_1, \theta_2, ..., \theta_K)$, where $\theta_i = \sigma_i^2 > 0$, $N = n_1 + n_2 + ... + n_K$ and $n_i / N \to \lambda_i \in (0,1)$, $i = 1, \ldots, K$.

The principles needed to prove the consistency of MLE's for (2.14), $K \geq 2$ are a little different from those for the Behrens - Fisher Problem, as follows:

(1) Show that a.e each of the diagonal terms of $\ddot{l}_N(\underline{\bar{\vartheta}}) \to -\infty$ for all sufficiently small $\delta > 0$.

(2) Roughly speaking, use the diagonal terms to control the off-diagonal terms as $\delta \to 0$.

Note that $(\mu_i - \mu_{i0})^2 > 0$, $i = 2,...,K$, $(\theta_j - \theta_{j0})^2 > 0$, $j = 1,...,K$ and the off-diagonal terms $(\mu_i - \mu_{i0})(\mu_j - \mu_{j0}), \ldots,$ and $(\mu_i - \mu_{i0})(\theta_j - \theta_{j0})$ are 'small'. The loglikelihood and its gradient vector under $H_0$ are given by

$$l_N(\underline{\vartheta}) = -\sum_{i=1}^{K} \frac{n_i}{2} \log(2\pi\theta_i) - \sum_{i=2}^{K}\sum_{j=1}^{n_i}(x_{ij} - \mu_i)^2 / 2\theta_i - \sum_{j=1}^{n_1}(x_{1j} - [\zeta\sqrt{\sum l_i^2\theta_i} - \sum_{i=2}^{K} l_i\mu_i]/l_1)^2 / 2\theta_1 ,$$

$$\frac{\partial l_N(\underline{\vartheta})}{\partial \mu_i} = \frac{\sum_{j=1}^{n_i}(x_{ij} - \mu_i)}{\theta_i} - \frac{l_i}{l_1\theta_1}\sum_{j=1}^{n_1}(x_{1j} - [\zeta\sqrt{\sum l_i^2\theta_i} - \sum_{i=2}^{K} l_i\mu_i]/l_1), \quad i = 2,...,K ,$$

$$\frac{\partial l_N(\underline{\vartheta})}{\partial \theta_1} = -\frac{n_1}{2\theta_1} + \frac{\sum_{j=1}^{n_1}(x_{1j} - [\zeta\sqrt{\sum l_i^2\theta_i} - \sum_{i=2}^{K} l_i\mu_i]/l_1)^2}{2(\theta_1)^2} + \frac{l_1\zeta\sum_{j=1}^{n_1}(x_{1j} - [\zeta\sqrt{\sum l_i^2\theta_i} - \sum_{i=2}^{K} l_i\mu_i]/l_1)}{2\theta_1\sqrt{\sum l_i^2\theta_i}} ,$$

$$\frac{\partial l_N(\underline{\vartheta})}{\partial \theta_i} = -\frac{n_i}{2\theta_i} + \frac{\sum_{j=1}^{n_i}(x_{ij} - \mu_i)^2}{2(\theta_i)^2} + \frac{l_i^2\zeta\sum_{j=1}^{n_1}(x_{1j} - [\zeta\sqrt{\sum l_i^2\theta_i} - \sum_{i=2}^{K} l_i\mu_i]/l_1)}{2l_1\theta_1\sqrt{\sum l_i^2\theta_i}}, \ i = 2,...,K .$$

(1) Show that a.e each of the diagonal terms of $\ddot{l}_N(\underline{\vartheta}) \to -\infty$ for all sufficiently small $\delta > 0$.

$$\frac{\partial^2 l_N(\overline{\underline{\vartheta}})}{(\partial\mu_i)^2} = -\frac{n_i}{\overline{\theta}_i} - \frac{l_i^2 n_1}{l_1^2\overline{\theta}_i} < 0\,, \quad i = 2,\ldots,K \text{ , satisfies (1).}$$

$$\frac{\partial^2 l_N(\overline{\underline{\vartheta}})}{(\partial\theta_1)^2} = \frac{n_1}{2(\overline{\theta}_1)^2} - \frac{\sum_{j=1}^{n_1}\left(x_{1j} - [\zeta\sqrt{\sum l_i^2\overline{\theta}_i} - \sum_{i=2}^{K} l_i\overline{\mu}_i]/l_1\right)^2}{(\overline{\theta}_1)^3} - \frac{n_1 l_1^2\zeta^2}{4\overline{\theta}_1(\sum l_i^2\overline{\theta}_i)}$$

$$-\left(\frac{l_1^3\zeta}{4\overline{\theta}_1(\sum l_i^2\overline{\theta}_i)^{3/2}} + \frac{l_1\zeta}{(\overline{\theta}_1)^2\sqrt{\sum l_i^2\overline{\theta}_i}}\right)\sum_{j=1}^{n_1}(x_{1j} - [\zeta\sqrt{\sum l_i^2\overline{\theta}_i} - \sum_{i=2}^{K} l_i\overline{\mu}_i]/l_1)$$

$$= \frac{n_1}{2(\overline{\theta}_1)^3}\left[\overline{\theta}_1 - 2\theta_{10} - 2(\mu_{10} - \overline{\mu}_1)^2 + o(1)\right] + \frac{n_1}{\overline{\theta}_1}\left[\frac{-l_1^2\zeta^2}{4(\sum l_i^2\overline{\theta}_i)} + o_{n_1}(1) + o_\delta(1)\right]$$

$$\leq \frac{4n_1}{27\theta_{10}^3}[-0.5\theta_{10} + 2\delta^2 + o(1)] + \frac{n_1}{\overline{\theta}_1}\left[\frac{-l_1^2\zeta^2}{4(\sum l_i^2\overline{\theta}_i)} + o_{n_1}(1) + o_\delta(1)\right]$$

$$\to -\infty.$$

$$\frac{\partial^2 l_N(\overline{\underline{\vartheta}})}{(\partial\theta_i)^2} = \frac{n_i}{2(\overline{\theta}_i)^2} - \frac{\sum_{j=1}^{n_i}(x_{ij} - \overline{\mu}_i)^2}{(\overline{\theta}_i)^3} - \frac{n_1 l_i^4\zeta^2}{4l_1^2\overline{\theta}_1(\sum l_i^2\overline{\theta}_i)}$$

$$-\left(\frac{l_i^4\zeta}{4l_1\overline{\theta}_1(\sum l_i^2\overline{\theta}_i)^{3/2}}\right)\sum_{j=1}^{n_1}(x_{1j} - [\zeta\sqrt{\sum l_i^2\overline{\theta}_i} - \sum_{i=2}^{K} l_i\overline{\mu}_i]/l_1)\,, \qquad i = 2,\ldots,K\,,$$

$$= \frac{n_i}{2(\overline{\theta}_i)^3}\left[\overline{\theta}_i - 2\theta_{i0} - 2(\mu_{i0} - \overline{\mu}_i)^2 + o(1)\right] + \frac{n_1}{\overline{\theta}_1}\left[\frac{-l_i^4\zeta^2}{4l_1^2(\sum l_i^2\overline{\theta}_i)} + o_{n_1}(1) + o_\delta(1)\right]$$

$$\leq \frac{4n_i}{27\theta_{i0}^3}[-0.5\theta_{i0} + 2\delta^2 + o(1)] + \frac{n_1}{\overline{\theta}_1}\left[\frac{-l_i^4\zeta^2}{4l_1^2(\sum l_i^2\overline{\theta}_i)} + o_{n_1}(1) + o_\delta(1)\right]$$

$$\to -\infty.$$

(2) Now, we will use diagonal terms to control the off-diagonal terms as $\delta \to 0$. We have that

$$\frac{\partial^2 l_N(\overline{\vartheta})}{\partial\mu_i\partial\mu_j} = -\frac{n_1 l_i l_j}{\overline{\theta}_1 l_1^2}, \quad i,j = 2,\ldots,K, \ i \neq j,$$

$$\frac{\partial^2 l_N(\overline{\vartheta})}{\partial\mu_i\partial\theta_1} = \frac{n_1 l_i \zeta}{2\overline{\theta}_1\sqrt{\sum l_m^2\overline{\theta}_m}} - \frac{l_i}{l_1(\overline{\theta}_1)^2}\sum_{j=1}^{n_1}(x_{1j} - [\zeta\sqrt{\sum l_m^2\overline{\theta}_m} - \sum_{i=2}^{K} l_i\overline{\mu}_i]/l_1), \quad i = 2,\ldots,K,$$

$$= \frac{n_1}{\overline{\theta}_1}\left[\frac{l_i\zeta}{2\sqrt{\sum l_m^2\overline{\theta}_m}} + o_{n_1}(1) + o_\delta(1)\right].$$

$$\frac{\partial^2 l_N(\overline{\vartheta})}{\partial\mu_i\partial\theta_i} = \frac{n_1 l_i^3\zeta}{2l_1^2\overline{\theta}_1\sqrt{\sum l_m^2\overline{\theta}_m}} - \frac{n_i(\overline{X}_i - \overline{\mu}_i)}{(\overline{\theta}_i)^2}, \quad i = 2,\ldots,K,$$

$$= \frac{n_1}{\overline{\theta}_1}\left[\frac{l_i^3\zeta}{2l_1^2\sqrt{\sum l_m^2\overline{\theta}_m}} + o_{n_i}(1) + o_\delta(1)\right]$$

$$\frac{\partial^2 l_N(\overline{\vartheta})}{\partial\mu_i\partial\theta_j} = \frac{n_1 l_i l_j^2\zeta}{\overline{\theta}_1 2l_1^2\sqrt{\sum l_m^2\overline{\theta}_m}}, \quad i,j = 2,\ldots,K, \ i \neq j.$$

$$\frac{\partial^2 l_N(\overline{\vartheta})}{\partial\theta_1\partial\theta_i} = -\frac{n_1 l_i^2\zeta^2}{4\overline{\theta}_1(\sum l_m^2\overline{\theta}_m)}$$
$$-\left(\frac{l_i^2 l_1\zeta}{4\overline{\theta}_1(\sum l_m^2\overline{\theta}_m)^{3/2}} + \frac{l_i^2\zeta}{2l_1(\overline{\theta}_1)^2\sqrt{\sum l_m^2\overline{\theta}_m}}\right)\sum_{j=1}^{n_1}(x_{1j} - [\zeta\sqrt{\sum l_m^2\overline{\theta}_m} - \sum_{i=2}^{K} l_i\overline{\mu}_i]/l_1)$$
$$= \frac{n_1}{\overline{\theta}_1}\left[\frac{-l_i^2\zeta^2}{4(\sum l_m^2\overline{\theta}_m)} + o_{n_1}(1) + o_\delta(1)\right], \quad i = 2,\ldots,K.$$

$$\frac{\partial^2 l_N(\overline{\vartheta})}{\partial\theta_i\partial\theta_j} = -\frac{n_1 l_i^2 l_j^2\zeta^2}{4l_1^2\overline{\theta}_1(\sum l_m^2\overline{\theta}_m)} - \frac{l_i^2 l_j^2\zeta}{4l_1\overline{\theta}_1\sqrt{\sum l_m^2\overline{\theta}_m}}\sum_{j=1}^{n_1}(x_{1j} - [\zeta\sqrt{\sum l_m^2\overline{\theta}_m} - \sum_{i=2}^{K} l_i\overline{\mu}_i]/l_1)$$

$$= \frac{n_1}{\overline{\theta}_1}\left[ -\frac{l_i^2 l_j^2 \zeta^2}{4 l_1^2 (\sum l_m^2 \overline{\theta}_m)} + o_{n_1}(1) + o_\delta(1) \right], \qquad i, j = 2,\ldots,K,\ \ i \neq j .$$

Finally, a Taylor expansion yields the expression

$$(\underline{\vartheta} - \underline{\vartheta}_0)^T \dot{l}_N(\underline{\vartheta}) = (\underline{\vartheta} - \underline{\vartheta}_0)^T \dot{l}_N(\underline{\vartheta}_0) + (\underline{\vartheta} - \underline{\vartheta}_0)^T \ddot{l}_N(\overline{\underline{\vartheta}})(\underline{\vartheta} - \underline{\vartheta}_0)$$

$$= \sum_{i=2}^{K} (\mu_i - \mu_{i0}) \frac{\partial l_N(\underline{\vartheta}_0)}{\partial \mu_i} + \sum_{j=1}^{K} (\theta_j - \theta_{j0}) \frac{\partial l_N(\underline{\vartheta}_0)}{\partial \theta_j}$$

$$+ \sum_{i=2}^{K} (\mu_i - \mu_{i0})^2 \frac{\partial^2 l_N(\overline{\underline{\vartheta}})}{(\partial \mu_i)^2} + \sum_{j=1}^{K} (\theta_j - \theta_{j0})^2 \frac{\partial^2 l_N(\overline{\underline{\vartheta}})}{(\partial \theta_j)^2}$$

$$+ \sum_{\substack{i=2,\ldots,K;\\ j=1,\ldots,K;\\ i\neq j}} (\mu_i - \mu_{i0})(\mu_j - \mu_{j0}) \frac{\partial^2 l_N(\overline{\underline{\vartheta}})}{\partial \mu_i \partial \mu_j} + \sum_{i=2}^{K}\sum_{j=1}^{K} (\mu_i - \mu_{i0})(\theta_j - \theta_{j0}) \frac{\partial^2 l_N(\overline{\underline{\vartheta}})}{\partial \mu_i \partial \theta_j}$$

$$+ \sum_{\substack{i,j=1,\ldots,K;\\ i\neq j}} (\theta_i - \theta_{i0})(\theta_j - \theta_{j0}) \frac{\partial^2 l_N(\overline{\underline{\vartheta}})}{\partial \theta_i \partial \theta_j} .$$

Let $A_i = (\mu_i - \mu_{i0})\frac{l_i}{l_1}$, for $i = 2,\ldots,K$, and $B_j = (\theta_j - \theta_{j0})\frac{l_j^2 \zeta}{2 l_1 \sqrt{\sum l_m^2 \overline{\theta}_m}}$, for $j = 1,\ldots,K$.

Then, we have:

$$(\underline{\vartheta} - \underline{\vartheta}_0)^T \dot{l}_N(\underline{\vartheta}) = (\underline{\vartheta} - \underline{\vartheta}_0)^T \dot{l}_N(\underline{\vartheta}_0) + (\underline{\vartheta} - \underline{\vartheta}_0)^T \ddot{l}_N(\overline{\underline{\vartheta}})(\underline{\vartheta} - \underline{\vartheta}_0)$$

$$\leq N\Bigg[ -\delta^2 \min\{\frac{2\lambda_2}{\theta_{20}},\ldots,\frac{2\lambda_K}{\theta_{K0}}\ \frac{4\lambda_1}{27\theta_{10}^3}[1.5\theta_{10} - 2\delta^2 + o(1)],\ \frac{4\lambda_K}{27\theta_{K0}^3}[1.5\theta_{K0} - 2\delta^2 + o(1)]\}$$

$$- \frac{\lambda}{\overline{\theta}_1}(\sum_{i=2}^{K} A_i - \sum_{j=1}^{K} B_j)^2 + o_N(1) + o_\delta(1) \Bigg]$$

$\to -\infty$, as $N \to \infty$, and as $\delta \to 0$.

where $o_N(1) \to 0$ as $N \to \infty$; $o_\delta(1) \to 0$ as $\delta \to 0$, which completes the verification.

## 2.5 An Iterative Method of Finding the MLE's under $H_0$ in (2.4)

- **Algorithm for Finding Candidate MLE's for a specific value of $\pi$**

Algorithm 2.1: Let $\underline{\hat{\mu}}^{(t)} = \{\hat{\mu}_1^{(t)}, \hat{\mu}_2^{(t)}, ..., \hat{\mu}_K^{(t)}\}$ and $\underline{\hat{\sigma}}^{2(t)} = \{\hat{\sigma}_1^{2(t)}, \hat{\sigma}_2^{2(t)}, ..., \hat{\sigma}_K^{2(t)}\}$ denote the candidate mle's under the null hypothesis at the t$^{th}$ step.

1. Start the iteration with the multiple initial points for $\underline{\hat{\mu}}^{(0)}$: $\underline{\hat{\mu}}^{(0)} = \underline{\bar{x}}$ (sample means); then from equation (2.20) calculate $\underline{\hat{\sigma}}^{2(0)}$.
2. Plug the value of $\underline{\hat{\sigma}}^{2(0)}$ into equation (2.19). Then, calculate $\underline{\hat{\mu}}^{(1)}$ and $\text{diff}(\underline{\hat{\mu}})^{(1)} = \left|\underline{\hat{\mu}}^{(1)} - \underline{\hat{\mu}}^{(0)}\right|$. Next, plug the value of $\underline{\hat{\mu}}^{(1)}$ into equation (2.20) to calculate $\underline{\hat{\sigma}}^{2(1)}$ and $\text{diff}(\underline{\hat{\sigma}}^{2})^{(1)} = \left|\underline{\hat{\sigma}}^{2(1)} - \underline{\hat{\sigma}}^{2(0)}\right|$.
3. Iteration: compute ($\underline{\hat{\mu}}^{(t)}, \underline{\hat{\sigma}}^{2(t)}$) from the previous ($\underline{\hat{\mu}}^{(t-1)}, \underline{\hat{\sigma}}^{2(t-1)}$) using equations (2.19) and (2.20) until the maximum of $\text{diff}(\underline{\hat{\mu}})^{(t)}$ and the maximum of $\text{diff}(\underline{\hat{\sigma}}^{2})^{(t)}$ are less than a very small value.
4. Use ($\underline{\hat{\mu}}^{(t)}, \underline{\hat{\sigma}}^{2(t)}$) as the MLE ($\underline{\hat{\mu}}, \underline{\hat{\sigma}}^{2}$) under $H_0$.

Preliminary simulations presented in Table 2.1 with $\{l_i, i = 1,2,3\}$= {-1/2, -1/2, 1} and $\pi$ =0.90 indicate that this iterative method appears to have good convergence rates, especially for large sample sizes where it always converged to a local maximum.

Table 2.1 Estimated Convergence Proportions Based on 10000 iterations

| X1, X2, X3 | Sample size (n1,n2,n3) | | | | |
|---|---|---|---|---|---|
| | (10,10,10) | (10,12,15) | (10,30,90) | (30,30,30) | (50,50,50) |
| N(0,1), N(0,1), N(1.56767, 1) | 0.8781 | 0.9417 | 0.9994 | 0.9877 | 0.9990 |
| N(0,1), N(0,10), N(3, 2.743164) | 0.8815 | 0.9372 | 0.9991 | 0.9962 | 0.9999 |

Furthermore, I carried out some simulations to investigate the performance of the estimators obtained from the algorithm given (2.5) in terms of BIAS and mean squared error (MSE).

- **Estimated Mean Squared Error of MLE's**

Although we do not have an explicit formula for the mean squared errors of the maximum likelihood estimators described in (2.5), denoted MSE(MLE's), I estimated them by simulation, as follows. First, I specified some values for sample sizes: $n$ = 10 (small), 30(medium), 100 (large), and parameters $K$=3, and $\{(\mu_i, \sigma_i^2), i = 1,2,\ldots,K$. All parameter values are listed in the third column in each table. Then, I generated 10000 independent data sets from each setting and computed W($\hat{\theta}$) = ($\hat{\theta}$ - $\theta$)$^2$, where $\theta = \mu_i$ or $\sigma_i^2$, $i$ = 1,2,…, $K$, for each data set. Those W($\hat{\theta}$) do not include the values where the algorithm did not converge. The mean of the resulting W's, denoted, $\overline{\mathrm{W}} = MS\hat{E}(\hat{\theta})$, is a consistent, unbiased estimator of MSE($\hat{\theta}$). The results are given in Table 2.2 below, where I report $\sqrt{MS\hat{E}(\hat{\mu}_i)}$ and $\sqrt{MS\hat{E}(\hat{\sigma}_i^2)}/\sigma_i^2$, the latter relative value since there is considerable variation in the $\sigma_i^2$'s in the study, $i = 1,\ldots,K$.

- **Estimated BIAS of MLE's**

I used the same approach and data sets as described above to estimate the biases of the maximum likelihood estimators , where now V($\hat{\theta}$) = ($\hat{\theta}$ - $\theta$), and $\theta = \mu_i$ or $\sigma_i^2$, $i$ = 1,2,…, $K$, for each data set. The mean of the resulting V($\hat{\theta}$)'s over data sets where the algorithm converged, denoted $\overline{\mathrm{V}}(\hat{\theta}) = BIA\hat{S}(\hat{\theta})$, is a consistent, unbiased estimator of $BIAS(\hat{\theta})$. The results are given in Table 2.3. The parameters and sample sizes setting are the same as those for estimated MSE. All parameter values are listed in the third column in each table.

Table 2.2 Estimated Relative MSE of the MLE's Based on 10000 iterations

Table 2.2.1 *K*=3, (a) $(\mu_2) = (0.5)$

| $\pi$ | $i$ | *Parameter Values* | | small sample sizes | | medium sample sizes | | large sample sizes | |
|---|---|---|---|---|---|---|---|---|---|
| | | $\mu_i$ | $\hat{\sigma}_i^2$ | $\hat{\mu}_i$ | $\hat{\sigma}_i^2$ | $\hat{\mu}_i$ | $\hat{\sigma}_i^2$ | $\hat{\mu}_i$ | $\hat{\sigma}_i^2$ |
| *0.55* | *1* | 0 | 1 | 0.2931 | 0.4501 | 0.1735 | 0.2581 | 0.0883 | 0.1411 |
| | *2* | 0.5 | 1 | 0.2417 | 0.3635 | 0.1364 | 0.1985 | 0.0768 | 0.1149 |
| | *3* | 0.4039 | 1 | 0.1720 | 0.4167 | 0.1054 | 0.3205 | 0.0520 | 0.1273 |
| *0.65* | *1* | 0 | 1 | 0.2919 | 0.4486 | 0.1726 | 0.2563 | 0.0894 | 0.1411 |
| | *2* | 0.5 | 1 | 0.2470 | 0.3644 | 0.1375 | 0.2013 | 0.0762 | 0.1149 |
| | *3* | 0.7219 | 1 | 0.1775 | 0.4101 | 0.1109 | 0.3106 | 0.0548 | 0.1281 |
| *0.75* | *1* | 0 | 1 | 0.2927 | 0.4564 | 0.1744 | 0.2583 | 0.0906 | 0.1404 |
| | *2* | 0.5 | 1 | 0.2464 | 0.3648 | 0.1375 | 0.1987 | 0.0762 | 0.1158 |
| | *3* | 1.0761 | 1 | 0.1871 | 0.3956 | 0.1233 | 0.2965 | 0.0583 | 0.1237 |
| *0.85* | *1* | 0 | 1 | 0.3012 | 0.4599 | 0.1746 | 0.2567 | 0.0917 | 0.1421 |
| | *2* | 0.5 | 1 | 0.2526 | 0.3692 | 0.1364 | 0.2010 | 0.0781 | 0.1166 |
| | *3* | 1.5194 | 1 | 0.2045 | 0.3775 | 0.1411 | 0.2809 | 0.0625 | 0.1170 |

Table 2.2.2 *K*=3, (b) $(\mu_2, \mu_3) = (0.5, 2)$

| $\pi$ | $i$ | *Parameter Values* | | small sample sizes | | medium sample sizes | | large sample sizes | |
|---|---|---|---|---|---|---|---|---|---|
| | | $\mu_i$ | $\hat{\sigma}_i^2$ | $\hat{\mu}_i$ | $\hat{\sigma}_i^2$ | $\hat{\mu}_i$ | $\hat{\sigma}_i^2$ | $\hat{\mu}_i$ | $\hat{\sigma}_i^2$ |
| *0.65* | *1* | 0 | 1 | 0.3191 | 0.4417 | 0.1811 | 0.2567 | 0.0990 | 0.1404 |
| | *2* | 0.5 | 1.5 | 0.3165 | 0.3595 | 0.1738 | 0.1999 | 0.1 | 0.1157 |
| | *3* | 2 | 20.0018 | 0.3971 | 0.3933 | 0.2851 | 0.3085 | 0.1233 | 0.1247 |
| *0.75* | *1* | 0 | 1 | 0.3159 | 0.4395 | 0.1794 | 0.2575 | 0.0985 | 0.1411 |
| | *2* | 0.5 | 1.5 | 0.3116 | 0.3602 | 0.1706 | 0.2 | 0.0990 | 0.1153 |
| | *3* | 2 | 6.10671 | 0.3490 | 0.3733 | 0.2498 | 0.2915 | 0.11 | 0.1186 |
| *0.85* | *1* | 0 | 1 | 0.3106 | 0.4578 | 0.1797 | 0.255 | 0.0954 | 0.1411 |
| | *2* | 0.5 | 1.5 | 0.3059 | 0.3601 | 0.1712 | 0.1969 | 0.0959 | 0.1145 |
| | *3* | 2 | 1.4920 | 0.2795 | 0.5249 | 0.2040 | 0.4087 | 0.0883 | 0.1686 |
| *0.95* | *1* | 0 | 1 | 0.2898 | 0.4278 | 0.1729 | 0.2587 | 0.09 | 0.1378 |
| | *2* | 0.5 | 1.5 | 0.2895 | 0.3475 | 0.1637 | 0.1959 | 0.09 | 0.1105 |
| | *3* | 2 | 0.5069 | 0.1735 | 0.3563 | 0.1237 | 0.2713 | 0.0548 | 0.1168 |

Table 2.3 Estimated BIAS of the MLE's Based on 10000 iterations

Table 2.3.1 $K$=3, (a) $(\mu_2) = (0.5)$

| $\pi$ | $i$ | Parameter Values | | small sample sizes | | medium sample sizes | | large sample sizes | |
|---|---|---|---|---|---|---|---|---|---|
| | | $\mu_i$ | $\hat{\sigma}_i^2$ | $\hat{\mu}_i$ | $\hat{\sigma}_i^2$ | $\hat{\mu}_i$ | $\hat{\sigma}_i^2$ | $\hat{\mu}_i$ | $\hat{\sigma}_i^2$ |
| 0.55 | 1 | 0 | 1 | 0.00008 | -0.0805 | -0.0006 | -0.0284 | 5.200e-04 | -0.0075 |
| | 2 | 0.5 | 1 | 0.0050 | -0.0581 | -0.0015 | -0.0186 | 4.805e-04 | -0.0047 |
| | 3 | 0.4039 | 1 | -0.0022 | -0.0247 | -0.0029 | -0.0061 | -2.534e-05 | -0.0041 |
| 0.65 | 1 | 0 | 1 | 0.0041 | -0.0769 | 0.0005 | -0.0254 | 0.0006 | -0.0080 |
| | 2 | 0.5 | 1 | 0.0101 | -0.0498 | 0.0005 | -0.0165 | 0.0008 | -0.0073 |
| | 3 | 0.7219 | 1 | -0.0078 | -0.0308 | -0.0052 | -0.0091 | -0.0008 | -0.0023 |
| 0.75 | 1 | 0 | 1 | 0.0083 | -0.0780 | 0.0039 | -0.03369 | 0.0013 | -0.0096 |
| | 2 | 0.5 | 1 | 0.0022 | -0.0562 | 0.00179 | -0.0166 | 0.0020 | -0.0036 |
| | 3 | 1.0761 | 1 | -0.0185 | -0.0237 | -0.00809 | -0.0111 | -0.0007 | -0.0025 |
| 0.85 | 1 | 0 | 1 | 0.0158 | -0.0728 | 0.0037 | -0.0255 | 0.0029 | -0.0073 |
| | 2 | 0.5 | 1 | 0.0111 | -0.0488 | 0.0031 | -0.0163 | 0.0005 | -0.0052 |
| | 3 | 1.5194 | 1 | -0.0220 | -0.0271 | -0.0114 | -0.0106 | -0.0017 | -0.0024 |

Table 2.3.2 $K$=3, (b) $(\mu_2, \mu_3) = (0.5, 2)$

| $\pi$ | $i$ | Parameter Values | | small sample sizes | | medium sample sizes | | large sample sizes | |
|---|---|---|---|---|---|---|---|---|---|
| | | $\mu_i$ | $\hat{\sigma}_i^2$ | $\hat{\mu}_i$ | $\hat{\sigma}_i^2$ | $\hat{\mu}_i$ | $\hat{\sigma}_i^2$ | $\hat{\mu}_i$ | $\hat{\sigma}_i^2$ |
| 0.65 | 1 | 0 | 1 | -0.0005 | -0.1000 | 0.0015 | -0.0324 | -0.0007 | -0.0100 |
| | 2 | 0.5 | 1.5 | 0.0081 | -0.0965 | -0.0042 | -0.0348 | 0.0004 | -0.0119 |
| | 3 | 2 | 20.0018 | -0.0393 | -0.2042 | -0.0211 | 0.0070 | -0.0043 | -0.0176 |
| 0.75 | 1 | 0 | 1 | 0.0018 | -0.0987 | 0.0020 | -0.0318 | 0.0018 | -0.0081 |
| | 2 | 0.5 | 1.5 | 0.0079 | -0.0925 | 0.0018 | -0.0302 | 0.0010 | -0.0109 |
| | 3 | 2 | 6.10671 | -0.0359 | -0.0700 | -0.0179 | -0.0199 | -0.0026 | -0.0066 |
| 0.85 | 1 | 0 | 1 | 0.0131 | -0.0696 | 0.00218 | -0.0316 | 0.0019 | -0.0097 |
| | 2 | 0.5 | 1.5 | 0.0203 | -0.0634 | 0.0028 | -0.0302 | -0.0002 | -0.0110 |
| | 3 | 2 | 1.4920 | -0.0339 | -0.0740 | -0.0168 | -0.0140 | -0.0041 | -0.0051 |
| 0.95 | 1 | 0 | 1 | 0.0268 | -0.0527 | 0.0104 | -0.0094 | 0.0047 | -0.0075 |
| | 2 | 0.5 | 1.5 | 0.0244 | -0.0266 | 0.0102 | -0.0069 | 0.0038 | -0.0044 |
| | 3 | 2 | 0.5069 | -0.0198 | -0.0271 | -0.0102 | -0.0171 | -0.0004 | -0.0018 |

The small entries in Table 2.2 indicate that the Algorithm 2.1 given in section 2.5 provides estimators that are close to being unbiased for the parameter values used in this study. As expected, bias tends to decrease as sample size increases. However, the estimated relative, root-mean-squared errors are only as low as the 10% range for the large sample sizes.

## 2.6 Two Methods of Computing a LRT Statistic Under $H_0$ in (2.4)

My null hypothesis is composite and I used two different methods to obtain LRT statistics. In both cases the algorithm defined above was used to estimate the maximum likelihood estimators:

<u>Method 1</u>: Compute the LRT statistic with the numerator obtained only for $\pi = \pi_0$, so that

$n_\pi = \Phi^{-1}(\pi_0)$.

<u>Method 2</u>: Select representative proportions $\{\pi_j;\ 0.5 < \pi_j \le \pi_0,\ j = 1,2,...,M;\ \pi_M = \pi_0\}$ and use $L_{N,M} = Max\{L_N(\underline{\mu}_j, \underline{\sigma}_j)\}$ in the numerator of the likelihood ratio statistic. Specifically, use the statistic

$$\Lambda = \frac{L_{N,M}}{L(\hat{\underline{\mu}}, \hat{\underline{\sigma}}^2)}.$$

As will be seen in Chapter 5, Method 1 does not work well, in the sense that it results in a test that need not be unbiased. Method 2 does perform well overall and appears to result in an unbiased test. Therefore, I will use Method 2 to investigate the parametric bootstrap used to calibrate the LRT in the full simulation study presented in Chapter 5. The general scheme for using a bootstrap to estimate a p-value using a statistic '$T$', where large values favor $H_1$ over $H_0$, is given below. I call this the *parametric bootstrap likelihood ratio* (*PBL*) test when $T$ is the LRT statistic.

Having observed $T = t$, a parametric bootstrap test is carried out as follows.

1. Obtain the mle $\hat{\underline{\theta}}_{0,N}$ (using Algorithm 2.1) constrained by $H_0$, where $\pi = \pi_0$.
2. Generate $R$ independent samples $\{\underline{x}^*_r\}$ from the model $P_{\hat{\theta}_{0,N}}$.
3. Calculate the value of the test statistics $T(\underline{x}^*_r)$ (using Method 2) for each resample.
4. Estimate the p-value by

$$\hat{p}^* = \frac{1 + \sum_{r=1}^{R} I(T(\underline{x}^*_r) \geq t))}{R + 1}.$$

A full simulation study of this method applied to my tests is given in Chapter 5.

## 2.7. Tests Based on Bootstrap Confidence Sets

An alternative method for testing the hypotheses in (1.2), which, as the reader will recall, amounts to choosing between ‘ IS($\underline{F}$) $\leq \Psi$ ’ and ‘IS($\underline{F}$) $> \Psi$ ’, is to use a bootstrap, parametric or nonparametric, to construct a one-sided lower confidence set $I$ for IS($\underline{F}$) and conclude that ‘IS($\underline{F}$) $> \Psi$ ’ if $I$ does not contain $\Psi$. The nonparametric bootstrap can be carried out by independently resampling from the data from each distribution and using the percentile method or the BCa (Bias Corrected and Accelerated) of constructing confidence sets. Hall and Martin (1988) applied this approach to the Behrens-Fisher problem. In section 4.2.3 I construct confidence sets by using a prepivoted bootstrap.

# Chapter 3 Average P-Values

As stated above, my inference problems lie within the general framework of constructing tests in the presence of nuisance parameters. The Behrens-Fisher problem is a famous illustration of how difficult this can be. In cases when a test statistic is pivotal under $H_0$, reporting a p-value and rejecting the null hypothesis if $p \leq \alpha$ provides an exact size $\alpha$ test. Consider, for example testing $H_0 : \mu = \mu_0$ vs $H_a : \mu > \mu_0$ based on a random sample $\underline{x}$ of size $n$ from $N(\mu, \sigma^2)$, with both parameters unknown. Here, $T$ = $\sqrt{n}(\bar{X} - \mu_0)/S$ has a t-distribution with $n$-1 degrees of freedom under $H_0$ for all values of the nuisance parameter $\sigma$ > 0. Note that using the concept of monotone likelihood ratio, this procedure still works if the null hypothesis is generalized to $H_0 : \mu \leq \mu_0$. Since an appropriate pivotal does not exist for the Behrens-Fisher problem, a special case of my tests, p-values that are uniformly distributed under $H_0$, may not be available. Instead, I will construct p-values to weigh the evidence in the data against $H_0$ by averaging p-values obtained as functions of nuisance parameters over a distribution on these unknown quantities, a procedure recently studied in other cases by Bayarri and Berger (2000.). In Section 3.4, I prove that average p-values are asymptotically 'correct' under mild conditions.

Suppose that an observable random variable $\underline{X}$ has a family of distributions indexed by a vector of parameters $\underline{\theta} = (\underline{\theta}_1, \underline{\theta}_2)$ and we want to test $H_0 : \underline{\theta}_1 = \underline{\theta}_{10}$, viewing $\underline{\theta}_2$ as a nuisance parameter. We assume that there is a test function $T(\underline{X}, \underline{\theta})$ such that $T(\underline{X}, (\underline{\theta}_{10}, \underline{\theta}_2))$ is pivotal for all $\underline{\theta}_2$ and having observed $\underline{X} = \underline{x}$, large values of $T(\underline{x}, (\underline{\theta}_{10}, \underline{\theta}_2)) \equiv t(\underline{x}, \underline{\theta}_2)$ support the alternative hypothesis $H_1$ over $H_0$. If $\underline{\theta}_2$ were known, a p-value, uniformly distributed under $H_0$, would be given by

$$p(\underline{\theta}_2) \equiv P(T(X, (\underline{\theta}_{10}, \underline{\theta}_2) \geq t(\underline{x}, \underline{\theta}_2)).$$

In the absence of knowledge as to the value of the nuisance parameter and following Barnard (1984), we can average $p(\underline{\theta}_2)$ over a distribution $P_2$ on $\underline{\theta}_2$, which may depend on the observed value of $\underline{X}$, yielding an *extended* p-value given by

$$\overline{p} = \int p(\underline{\theta}_2) P_2(d\underline{\theta}_2 \mid \underline{x}), \tag{3.1}$$

which although not necessarily uniformly distributed under the null hypothesis will ,as I will show, behaves like a true p-value for my problems, in some cases, in the sense that rejecting $H_0$ when $\overline{p} \leq \alpha$ leads asymptotically to an approximate size $\alpha$ test. I will develop and explore this procedure in this chapter, where having observed $X = x$, $P_2$ is a Fiducial distribution on $\underline{\theta}_2$ and when $P_2$ is a posterior distribution on $\underline{\theta}_2$. Also, both distributions can be used to compute the 'probability' that the null hypothesis is true given the observed data. The average p-values studied in Bayarri and Berger (2000) are similar in persepective to (3.1) but different in structure. They consider the situation where the test function is actually a *statistic* ,$T(\underline{X})$, whose distribution ,$F_T(\cdot \mid \underline{\theta})$, depends on an unknown parameter $\underline{\theta}$. They then define an *average* p-value as

$$\text{p} = \int (1 - F_T(T(\underline{x}) \mid \underline{\theta}))\pi(d\underline{\theta} \mid \underline{x}),$$

where $\pi(d\underline{\theta} \mid \underline{x})$ is a distribution on $\underline{\theta}$ which may depend on having observed $\underline{X} = \underline{x}$. This p-value is the same as the one given in (3.1) in some cases.

Tsui and Weerahandi (1989) have proposed another extended p-value, called a *generalized p-value*, to deal with nuisance parameters, which I will investigate in future research.

### 3.1 Fiducial P-Value

The concept of Fiducial probability leads to a form of statistical inference based on inverse probability without requiring prior probability distributions. It was first proposed by E.B. Wilson (1927), and then developed by R. A. Fisher (1956). Edwards (1997) is an informative essay on this mode of inference. However, Fiducial inference is not widely used now and is not always mathematically consistent and free of contradictions in the multivariate case. I will use the symbol $FP$ to denote Fiducial probabilities.

In the two-sample case, $K$= 2, $l = (1,-1)^T$ and $\pi$ = 0.5, equation (2.1) is the Behrens-Fisher problem. Barnard (1984) shows that the Behrens-Fisher approach (which can be viewed as Fiducial inference) compares favorably with Welch's test. In this chapter, I will extend the use of the Fiducial approach to test the hypothesis in equation (2.3), for at least three normal distributions.

Again, consider the hypotheses in (2.4), and for fixed $\underline{\sigma}^2$ the test statistic given (2.5), and repeated here for the reader's convenience,

$$T(\underline{\sigma}^2) = \frac{\sum l_i \overline{X}_i / \sqrt{\sum l_i^2 \sigma_i^2 / n_i}}{\sqrt{\dfrac{\sum (n_i - 1) S_i^2 / \sigma_i^2}{(N-K)}}} \tag{3.2}$$

$$= \frac{Z+\delta}{W} \sim t'(N-K, \delta = \sum l_i \mu_i / \sqrt{\sum l_i^2 \sigma_i^2 / n_i}) \ ,$$

where again $t'(N-K,\delta)$ denotes a non-central $t$ distribution with non-centrality parameter $\delta$ and $N-K$ degrees of freedom. Let $t_{obs}(\underline{\sigma}^2)$ be the observed value of $T(\underline{\sigma}^2)$. If $\{\sigma_i^2\}$ were known, a p-value could be defined by

$$p(\underline{\sigma}^2) = p-value(\underline{\sigma}^2) = P(\ T \geq t_{obs}(\underline{\sigma}^2)\ )$$

A Fiducial p-value is given by:

$$
\begin{aligned}
\overline{p}_{Fiducial} &= \int p(\underline{\sigma}^2) f(\underline{\sigma}^2 \mid data) d\underline{\sigma}^2 \\
&= \int p(\underline{\sigma}^2) F(d\underline{\sigma}^2 \mid data)
\end{aligned} \tag{3.3}
$$

where $f(\underline{\sigma}^2 \mid data)$ is a joint Fiducial density of $\underline{\sigma}^2$ based on the data, obtained as follows.

Since $\frac{(n_i - 1)S_i^2}{\sigma_i^2}$ has a chi-square distribution with degree of freedom $(n_i - 1)$, for *fixed* sample variance ($S_i^2$) and fixed sample size ($n_i$), the Fiducial distribution of $\sigma_i^2$ is given by:

$$
\sigma_i^2 \mid data \underset{FD}{\equiv} \frac{(n_i - 1)S_i^2}{U_i} \quad , \text{ for } i = 1,2,\ldots,K, \tag{3.4}
$$

$\{U_i \mid data \sim \chi^2(n_i - 1)\}$ and the distributions in (3.4) are then scaled-inverse chi-squares with scale factors $(n_i - 1)S_i^2$ and degrees of freedom $(n_i - 1)$, $i = 1,2,\ldots,K$. The hypotheses we are interested in are given in (2.4).

If we further assume that the K variances are independent, the Fiducial joint density ($f(\underline{\sigma}^2 \mid data)$) of $\underline{\sigma}^2$ based on the data is given by

$$
\begin{aligned}
f(\underline{\sigma}^2 \mid data) &= f(\sigma_1^2 \mid S_1^2,\, n_1) f(\sigma_2^2 \mid S_2^2, n_2) \ldots f(\sigma_K^2 \mid S_K^2,\, n_K) \\
&= \prod_{i=1}^{K} (f(\sigma_i^2 \mid S_i^2,\, n_i),
\end{aligned}
$$

where $f(\sigma_i^2 \mid S_i^2,\, n_i)$ is the Fiducial density function of $\sigma_i^2$ given by the data.

That is:

$$
\int_{\sigma_i^2}^{\infty} f(\sigma_i^2 \mid S_i^2, n_i)\, d\sigma_i^2 = \Pr\{\chi^2_{(n_i - 1)} < \frac{(n_i - 1)\, S_i^2}{\sigma_i^2}\}.
$$

Substituting these Fiducial distributions into (3.3) gives an explicit, complicated expression for $\overline{p}_{Fiducial}$. For the case of two-sample Behrens-Fisher problem, the Fiducial p-value (3.3) is a generalized p-value as given by Tsui and Weerahandi (1989). The Fiducial approach also leads to an evaluation of the 'probability' of $H_0$, as follows.

Since $$(\sum l_i \mu_i) \mid \underline{\sigma}^2, Data \sim N(\sum l_i \bar{x}_i, \sum l_i^2 \sigma_i^2 / n_i), \tag{3.5}$$

and under $H_0$,

$$\sum l_i \mu_i \le n_\pi \sqrt{\sum l_i^2 \sigma_i^2} \quad , \tag{3.6}$$

we have that

$$FP(H_0 \mid \underline{\sigma}^2, Data) = \Phi(\frac{n_\pi \sqrt{\sum l_i^2 \sigma_i^2} - \sum l_i \bar{x}_i}{\sqrt{\sum l_i^2 \sigma_i^2 / n_i}}) \ . \tag{3.7}$$

Averaging over the distributions in (3.4), yields

$$\begin{aligned} FP(H_0 \mid Data) &= E_{\underline{\sigma}^2}[P(H_0 \mid \underline{\sigma}^2, Data)] \\ &= E_{\underline{\sigma}^2}[\Phi(\frac{n_\pi \sqrt{\sum l_i^2 \sigma_i^2} - \sum l_i \bar{x}_i}{\sqrt{\sum l_i^2 \sigma_i^2 / n_i}})] \\ &= E_{\underline{U}}[\Phi(\frac{n_\pi \sqrt{\sum (n_i - 1) S_i^2 l_i^2 / U_i} - \sum l_i \bar{x}_i}{\sqrt{\sum \frac{(n_i - 1) S_i^2 l_i^2}{n_i U_i}}})] \end{aligned} \tag{3.8}$$

Note that $FP(H_0 \mid Data) \neq \bar{p}_{Fiducial}$ here, but they are equivalent using another simple test statistic $Z(\underline{\sigma}^2)$ given in (3.19) below. Small values of $FP(H_0 \mid Data)$ could be taken as evidence in support of the alternative hypothesis.

### 3.2 An Approximation to the Fiducial P-Value

The Fiducial p-value in (3.3) can in principle be computed by numerical integration. However, this can be difficult to carry out accurately. An approximation to the Fiducial p-value can be obtained by a Monte Carlo simulation, as follows. Instead of numerically approximating the complicated integral of (3.3), independently select a large number (B) values of $\sigma_i^2$ (say, $\sigma_{i(rep)}^2$) from the Fiducial distribution of $\sigma_i^2$ , $i = 1,2,\ldots,K$. Next, using those independent values $\sigma_{i(rep)}^2$ ($i = 1,\ldots,K$) calculate the value of $T(\underline{\sigma}_{(rep)}^2)$ for each replication. Then, evaluate the p-value, denoted by $\hat{p}(\underline{\sigma}_{(rep)}^2)$. The average of those p-values is an approximation of the Fiducial p-value, denoted by $\bar{\hat{p}}_{Fiducial}$ and given by

$$\bar{\hat{p}}_{Fiducial} = \sum_{rep=1}^{B} \hat{p}(\underline{\sigma}_{(rep)}^2) / B \quad , \tag{3.9}$$

where $\hat{p}(\underline{\sigma}_{(rep)}^2) = P(\,T \geq t_{obs}(\underline{\sigma}_{(rep)}^2))$ and $t_{obs}(\underline{\sigma}_{(rep)}^2)$ is the observed value of $T(\underline{\sigma}_{(rep)}^2)$. Here, we take $T$ to have a non-central t-distribution with non-centrality parameter $\delta$ = $\sqrt{n_{(k)}}\,(n_\pi)$ and $\nu = N - K$ degrees of freedom, which provides a conservative test.

This Fiducial p-value is a generalized p-value that will hopefully lead to a test of size at most or a little above $\alpha$ that also has good power. Based on a preliminary simulation (given in the Appendix A, Figure A.4), the distribution of the approximation to the Fiducial p-value in (3.9) appears, as desired, to be approximately uniform under the null hypothesis only if the sample sizes are equal and large enough ($n$ = 100). But for unequal sample sizes and for small equal sample sizes ($n$ = 10), the P-values have a highly skewed distribution under $H_0$. The unequal-sample-sizes case can be improved by using the replication estimates of the non-centrality parameter $\delta$ instead of an upper bound as described in (2.11). Again using

$$\delta = n_\pi \frac{\sqrt{\sum l_i^2 \sigma_i^2}}{\sqrt{\sum l_i^2 \sigma_i^2 / n_i}}, \tag{3.10}$$

plug the replication $\sigma^2_{i(rep)}$ into this equation in place of $\sigma^2_i$. Then, the test statistic $T(\underline{\sigma}^2)$ in equation (3.2) is treated as having a noncentral t-distribution with non-centrality parameter $\delta_{(rep)}$ and *N-K* degrees of freedom. Using

$$T(\underline{\sigma}^2_{(rep)}) \sim \text{t' } (N\text{-}K,\ \delta_{(rep)} = \frac{\sqrt{\sum l_i^2 \sigma^2_{i(rep)}}}{\sqrt{\sum l_i^2 \sigma^2_{i(rep)} / n_i}} n_\pi ) \ , \qquad (3.11)$$

results in an approximation of the Fiducial p-value given by:

$$\bar{\hat{p}}_{Fiducial} = \sum_{rep=1}^{B} \hat{p}(\underline{\sigma}^2_{(rep)}) / B \ , \qquad (3.12)$$

where $\hat{p}(\underline{\sigma}^2_{(rep)}) = P(T \geq t_{obs}(\underline{\sigma}^2_{(rep)}))$ , $t_{obs}(\underline{\sigma}^2_{(rep)})$ is an observed value of $T(\underline{\sigma}^2_{(rep)})$, and $\underline{\sigma}^2_{(rep)} = \{\sigma^2_{1(rep)},\ \sigma^2_{2(rep)}, \ldots,\ \sigma^2_{k(rep)}\}$, for rep =1,2,…,B. Each $\sigma^2_{i(rep)}$ has a scaled-inverse chi-square distribution with scale factor $(n_i - 1)S_i^2$ and degree of freedom $(n_i - 1)$ (in equation 3.4) and they are independent.

Preliminary simulations for Fiducial p-value in (3.12) (given in Appendix A, Figures A.5-A.6) indicate that this approach appears to yield Fiducial p-values that are approximately uniformly distributed under $H_0$ and tests that have good power, except for small sample sizes.

- **An Example of the Fiducial P-Value**

Assume we have three independent samples from normal distributions having sample means $\bar{x}_i$, sample size $n_i$, and sample variance $S_i^2$, respectively, *i* =1,2,3. Suppose we want to test:

$$H_0 : P(X_3 > \frac{(X_1 + X_2)}{2}) \leq \pi \ ,$$

$$H_a : P(X_3 > \frac{(X_1 + X_2)}{2}) > \pi \ , \qquad (3.13)$$

using the test statistic $T$ in equation (3.2) with $\{l_i\} = (-1/2, \quad -1/2, \quad 1)$.

Preliminary simulations for the Fiducial p-value in (3.9) and (3.12) are given in the Appendix A, Figures A.4-A.6. In Chapter 5, I conducted a full-scale simulation study to investigate the size and power of this test and compare the results to the frequentist tests given in Chapter 2.

### 3.3 Averaging Over a Posterior Distribution

In this section, we present a Bayesian approach to testing the Hypotheses in Equation (2.4). A conjugate prior distribution and a semi-conjugate prior distribution will be considered in my future research. Here, we use a Jeffrey's-type noninformative prior, given by

$$p(\underline{\mu}, \quad \underline{\sigma}^2) = c\frac{[\prod_{i=1}^{K} I_{(0,\infty)}(\sigma_i^2)]}{\prod_{i=1}^{K}\sigma_i^2} \quad , \tag{3.14}$$

where $c$ is an arbitrary positive constant.

The posterior distributions of $\underline{\mu}$ and $\underline{\sigma}^2$ are specified by

$$\mu_i \mid \sigma_i^2, \underline{x}_i \sim N(\bar{x}_i, \sigma_i^2 / n_i) \tag{3.15}$$

where $\{\mu_i \mid \sigma_i^2, \underline{x}_i\}_{i=1}^{K}$ are independent and

$$\sigma_i^2 \mid \underline{x}_i \underset{D}{\equiv} \frac{(n_i - 1)S_i^2}{U_i} \quad , \tag{3.16}$$

$\{U_i \mid \underline{x}_i \sim \chi^2(n_i - 1)\}$ and $\{\sigma_i^2 \mid \underline{x}_i\}$ are independent. These distributions can then be used in (3.1) to compute an average p-value. The distributions in (3.16) are scaled-inverse chi-square distributions with scale factors $(n_i - 1)S_i^2$ and degrees of freedom $(n_i - 1)$, $i$ =1,2,…, $K$. The hypotheses we are interested in are given in (2.4). Equations (3.5) and (3.6) imply that the posterior probability of $H_0$ given $\underline{\sigma}^2$ is

$$P(H_0 \mid \underline{\sigma}^2, Data) = \Phi(\frac{n_\pi \sqrt{\sum l_i^2 \sigma_i^2} - \sum l_i \bar{x}_i}{\sqrt{\sum l_i^2 \sigma_i^2 / n_i}}) . \quad (3.17)$$

Averaging over the distributions in (3.17), yields

$$\begin{aligned} P(H_0 \mid Data) &= E_{\underline{\sigma}^2}[P(H_0 \mid \underline{\sigma}^2, Data)] \\ &= E_{\underline{\sigma}^2}[\Phi(\frac{n_\pi \sqrt{\sum l_i^2 \sigma_i^2} - \sum l_i \bar{x}_i}{\sqrt{\sum l_i^2 \sigma_i^2 / n_i}})] \\ &= E_{\underline{U}}[\Phi(\frac{n_\pi \sqrt{\sum (n_i - 1) S_i^2 l_i^2 / U_i} - \sum l_i \bar{x}_i}{\sqrt{\sum \frac{(n_i - 1) S_i^2 l_i^2}{n_i U_i}}})] . \quad (3.18) \end{aligned}$$

Let us define another simple test statistic:

$$Z(\underline{\sigma}^2) = \frac{\sum l_i \bar{X}_i - n_\pi \sqrt{\sum l_i^2 \sigma_i^2}}{\sqrt{\sum l_i^2 \sigma_i^2 / n_i}} . \quad (3.19)$$

If $\underline{\sigma}^2$ were known, then the test statistic $Z(\underline{\sigma}^2)$ would have a standard normal distribution under the upper boundary of $H_0$ in (2.4). And a p-value for $H_0$ (2.4) could be defined by

$$p(\underline{\sigma}^2) = p - value(\underline{\sigma}^2) = P(Z \geq z_{obs}(\underline{\sigma}^2)), \quad (3.20)$$

where $Z \sim N(0,1)$; a posterior p-value could be defined by

$$\bar{p}_{post} = \int p(\underline{\sigma}^2) d(\underline{\sigma}^2 \mid Data)\, d\underline{\sigma}^2 , \quad (3.21)$$

where $D(\underline{\sigma}^2 \mid data)$ is a joint posterior distribution of $\underline{\sigma}^2$. Comparing (3.21) with equation (3.18), it is easy to see that $P(H_0 \mid Data) = \bar{p}_{post}$. I will use simulation to study the performance of tests based on treating $P(H_0 \mid Data)$ in (3.18) as a p-value for test statistic $Z(\underline{\sigma}^2)$ in (3.19) and make comparisons to my other tests.

There are many different techniques for simulating draws from complicated distributions, such as: Rejection Sampling, Gibbs Sampling, and the Metropolis algorithm. In my future research I will explore using these methods.

### 3.4 Consistency of "Average" P-Value

Here we show that the average p-values defined in (3.1) are asymptotically correct under mild conditions, a concept which does not appear to have yet been treated in the literature. Partition $\underline{\theta} = \{\underline{\theta}_1, \underline{\theta}_2\}$, where $\underline{\theta}_2 \in \Xi \subset \mathbb{R}^r$, $r \geq 1$, is viewed as a vector of *nuisance* parameters. Many of our tests have the form $H_0 : \zeta(\underline{\theta}) = \zeta_0$ vs $H_1 : \zeta(\underline{\theta}) > \zeta_0$, where $\zeta(\theta) = \zeta$ is a real valued function. Suppose that there is a real valued pivotal $T_N(\underline{X}_N, \zeta, \underline{\theta}_2)$ such that for a known function $H_N(.)$ of $\underline{x}_N$, $\zeta$ and $\underline{\theta}_2$ :

$$P_{\underline{\theta}}\ (T_N \leq T_N(\underline{x}_N, \zeta, \underline{\theta}_2)) \equiv H_N(\underline{x}_N \mid \zeta, \underline{\theta}_2)$$

$$\equiv 1\text{-}\ \bar{H}_N(\underline{x}_N \mid \zeta, \underline{\theta}_2)\,.$$

Based on data $\underline{X}_N = \underline{x}_N$, suppose that large values of $T_N(\underline{x}_N, \zeta_0, \underline{\theta}_2)$ support $H_1$ over $H_0$, whatever these hypotheses happen to be. Having observed $\underline{X}_N = \underline{x}_N$, $T_N(\underline{x}_N, \zeta_0, \underline{\theta}_2) \equiv t_{obs}(\underline{\theta}_2)$ and if $\underline{\theta}_2$ *were known*, a p-value uniformly distributed under $H_0$ when the data were generated by $\underline{\theta} = (\underline{\theta}_1, \underline{\theta}_2)$ with $\zeta(\underline{\theta}) = \zeta_0$, would be given by

$$p_N(t_{obs}(\underline{\theta}_2)) = P(T_N > t_{obs}(\underline{\theta}_2) \mid \zeta_0, \underline{\theta}_2) \qquad (3.22)$$

$$= \bar{H}_N(\underline{x}_N \mid \zeta_0, \underline{\theta}_2)$$

Let $\pi_N(\underline{\theta}_2 \mid \underline{x}_N)$ be a joint continuous (Posterior or Fiducial) density on $\Theta_2 = \{\underline{\theta}_2\}$, $N \geq 1$. Then, as defined above, an "average" p-value over $\pi_N(\underline{\theta}_2 \mid \underline{x}_N)$ is given by

$$\bar{p}_N(\underline{x}_N) = \int \bar{H}_N(\underline{x}_N \mid \zeta_0, \underline{\theta}_2)\pi_N(\underline{\theta}_2 \mid \underline{x}_N)d\underline{\theta}_2 . \qquad (3.23)$$

**Definition**: The average p-value ( $\bar{p}$ ) is *consistent* if for all $\{\underline{\theta}^* = (\underline{\theta}_1^*, \underline{\theta}_2^*),\ \zeta(\underline{\theta}^*) = \zeta_0\}$,

$$\left|\bar{p}_N(\underline{X}_N) - p_N(\underline{X}_N \mid \underline{\theta}_2^*)\right| \to 0 \quad \text{a.e. } P_{\underline{\theta}^*}, \quad \text{as } N \to \infty \qquad (3.24)$$

where $\underline{\theta}^* = (\underline{\theta}_1^*, \underline{\theta}_2^*)$ denotes the true parameter vector.

Since for all $N$, $p_N(\underline{X}_N \mid \underline{\theta}_2^*)$ is uniformly distributed when $\underline{X}_N \sim P_{\underline{\theta}^*}$, it follows from Slutsky's Theorem that a consistent $\{\bar{p}_N(\underline{X}_N)\}$ is asymptotically uniformly distributed with respect to $P_{\underline{\theta}^*}$. This makes $\bar{p}_N(\underline{X}_N)$ what Bayarri and Berger (2000) call a *frequentist p-value.*

**Theorem 2**: Suppose that for all $\underline{x}_N$ and $\zeta_0$, $\bar{H}_N(\underline{x}_N \mid \zeta_0, \underline{\theta}_2)$ is differentiable with respect to $\underline{\theta}_2$ and, as , $N \to \infty$ ,for a decreasing sequence of positive constants $\{\varepsilon_N\} \to 0$ , for all $\underline{\theta}^* = (\underline{\theta}_1^*, \underline{\theta}_2^*)$ for which $\zeta(\underline{\theta}^*) = \zeta_0$, a.e. $P_{\underline{\theta}^*}$

$$\text{(i) } P_{\underline{\theta}^*}(\left\|\underline{\theta}_2 - \underline{\theta}_2^*\right\| > \varepsilon_N \mid \underline{X}_N) \to 0 \ , \qquad (3.25)$$

$$\text{(ii) } \varepsilon_N \sum_{j=1}^{r} \sup_{\left\|\underline{\theta}_2 - \underline{\theta}_2^*\right\| \le \varepsilon_N} \left|h_N^{(j)}(\underline{X}_N \mid \zeta_0, \underline{\theta}_2)\right| \to 0 \ , \qquad (3.26)$$

where $\|.\|$ denotes a Euclidean distance and $h_N^{(j)}(\underline{X}_N \mid \zeta_0, \underline{\theta}_2) = \partial \bar{H}_N(\underline{X}_N \mid \zeta_0, \underline{\theta}_2)/\partial\theta_j$, $\theta_j$ in $\underline{\theta}_2$, are the partial derivatives of $\bar{H}_N$ at $\theta_j$, which are assumed to be jointly continuous functions of $(\underline{X}_N, \underline{\theta}_2)$, $N \ge 1$. Then,

$$\left|\bar{p}_N(\underline{X}_N) - p_N(\underline{X}_N \mid \underline{\theta}_2^*)\right| \to 0 \quad \text{a.e } P_{\underline{\theta}^*}, \quad \text{when } N \to \infty$$

**Proof**:

Let $B = \{\underline{\theta}_2 \ ; \left\|\underline{\theta}_2 - \underline{\theta}_2^*\right\| \le \varepsilon_N\}$, $b_N^{(j)}(\underline{x}_N) = \sup\{\left|h_N^{(j)}(\underline{x}_N \mid \zeta_0, \underline{\theta}_2)\right|; \underline{\theta}_2 \in B\}$, $a_j = (\underline{\theta}_2 - \underline{\theta}_2^*)_j$,

for $j$ =1,…, $r$. Then , for $\underline{\theta} = (\underline{\theta}_1^*, \underline{\theta}_2)$ , $\overline{\underline{\theta}}_2$ a point between $\underline{\theta}_2$ and $\underline{\theta}_2^*$, partitioning the parameter space and using a first-order Taylor expansion, we obtain:

$$
\begin{aligned}
&\left|\overline{p}_N(\underline{X}_N) - p_N(\underline{X}_N \mid \underline{\theta}_2^*)\right| \le \\
&\qquad \int_{\left\|\underline{\theta}_2 - \underline{\theta}_2^*\right\| \le \varepsilon_N} \left|\overline{H}_N(\underline{X}_N \mid \zeta_0, \underline{\theta}_2) - \overline{H}_N(\underline{X}_N \mid \zeta_0, \underline{\theta}_2^*)\right| \pi_N(\underline{\theta}_2 \mid \underline{X}_N)\, d\underline{\theta}_2 \\
&\qquad + \int_{\left\|\underline{\theta}_2 - \underline{\theta}_2^*\right\| > \varepsilon_N} \left|\overline{H}_N(\underline{X}_N \mid \zeta_0, \underline{\theta}_2) - \overline{H}_N(\underline{X}_N \mid \zeta_0, \underline{\theta}_2^*)\right| \pi_N(\underline{\theta}_2 \mid \underline{X}_N)\, d\underline{\theta}_2 \\
&\qquad \le \int_{\left\|\underline{\theta}_2 - \underline{\theta}_2^*\right\| \le \varepsilon_N} \left|(\underline{\theta}_2 - \underline{\theta}_2^*)' h_N(\underline{X}_N \mid \zeta_0, \overline{\underline{\theta}}_2)\right| \pi_N(\underline{\theta}_2 \mid \underline{X}_N) d\underline{\theta}_2 \\
&\qquad + \int_{\left\|\underline{\theta}_2 - \underline{\theta}_2^*\right\| > \varepsilon_N} \pi_N(\underline{\theta}_2 \mid \underline{X}_N)\, d\underline{\theta}_2 \\
&\qquad \le \int_{\left\|\underline{\theta}_2 - \underline{\theta}_2^*\right\| \le \varepsilon_N} \sum \left|a_j b_N^{(j)}(\underline{X}_N)\right| \pi_N(\underline{\theta}_2 \mid \underline{X}_N)\, d\underline{\theta}_2 + \int_{\left\|\underline{\theta}_2 - \underline{\theta}_2^*\right\| > \varepsilon_N} \pi_N(\underline{\theta}_2 \mid \underline{X}_N)\, d\underline{\theta}_2 \\
&\qquad \le \int_{\left\|\underline{\theta}_2 - \underline{\theta}_2^*\right\| \le \varepsilon_N} \varepsilon_{\mathrm{N}} \sum b_N^{(j)}(\underline{X}_N)\, \pi_N(\underline{\theta}_2 \mid \underline{X}_N)\, d\underline{\theta}_2 + \int_{\left\|\underline{\theta}_2 - \underline{\theta}_2^*\right\| > \varepsilon_N} \pi_N(\underline{\theta}_2 \mid \underline{X}_N)\, d\underline{\theta}_2 \\
&\qquad = \varepsilon_N \sum_{j=1}^{r} {}_{\underline{\theta}_2, \left\|\underline{\theta}_2 - \underline{\theta}_2^*\right\| \le \varepsilon_N} b_N^{(j)}(\underline{X}_N) + P_{\underline{\theta}^*}(\left\|\underline{\theta}_2 - \underline{\theta}_2^*\right\| > \varepsilon_N \mid \underline{X}_N) \\
&\qquad \to 0 \quad \text{a.e. as } N \to \infty,
\end{aligned}
$$

which completes the proof.

Berger(1985) gives conditions under which $\sqrt{N}(\underline{\theta}_2 - \underline{\theta}_2^* \mid \underline{x}_N)$ is almost surely, asymptotically normally distributed with a positive-definite covariance matrix, which can in some cases be used to verify (i) of Theorem 2. One need not subscribe to either a Bayesian or Fiducial approach to apply Theorem 2. For frequentists, where the data are generated from $P_{\underline{\theta}_1, \underline{\theta}_2^*}$ with fixed $(\underline{\theta}_1, \underline{\theta}_2^*)$, as long as the prior density is positive at $\underline{\theta}_2^*$, and (i) and (ii) hold, consistency prevails.

**Example 3.1**: Suppose $\{X_j\}_{j=1}^{N} \sim$ iid $N(\mu,\sigma^2)$ and we want to test $H_0 : \mu = \mu_0$, vs $H_1 : \mu > \mu_0$. We use this example even though the familiar t-test provides an alternative, exact solution, since it leads to a relatively easy illustration of Theorem 2. With $\theta_2 = \sigma^2$ and $T_N(\underline{x}_N,\sigma^2) = \sqrt{N}(\overline{x}_N - \mu_0)/\sigma$, $p(\underline{x}_N,\sigma^2) = \overline{H}_N(\underline{x}_N \mid \mu_0,\sigma^2) = \Phi[\sqrt{N}(\overline{x}_N - \mu_0)/\sigma)]$. Taking a Jeffreys-type noninformative prior of the form

$$p(\sigma^2) = (c/\sigma^2)\, I_{(0,\infty)}(\sigma^2) \;,$$

the posterior distribution of $\sigma^2$ is given by:

$$\pi(\sigma^2 \mid \underline{x}_N) = IG((N-1)/2,\; (N-1)S_N^2/2)\,, \tag{3.27}$$

an inverse gamma distribution corresponding to $(N-1)S_N^2/U$, where $U$ has a chi-square distribution with $N-1$ degrees of freedom. Take $\varepsilon_N = N^{-1/2+q}$, 0<$q$<1/2. Condition (i) holds since, using the asymptotic standard normality of the posterior distribution of $\sqrt{N}(\sigma^2 - \sigma^{2*})/\gamma$, where $\gamma^2$ is the limiting variance, $\phi$ denotes the standard normal density, a.e. $P_{\mu_{0\sigma^*}}$,

$$\begin{aligned} Lim P_{\mu_0\sigma^*}(\left|\sigma^2 - \sigma^{2*}\right| > \varepsilon_N \mid \underline{X}_N) &= Lim[2(1-\Phi(\sqrt{N}\varepsilon_N/\gamma))] \\ &= 0. \end{aligned} \tag{3.28}$$

To verify condition (ii), we have that

$$\begin{aligned} \left|h_N(T_N \mid \mu_0,\sigma^2)\right| &= \left|\frac{\sqrt{N}}{2(\sigma^2)^{3/2}}(\mu_0 - \overline{X}_N)\right| \phi((\mu_0 - \overline{X}_N)\sqrt{N}/\sigma) \\ &\le \frac{1}{2\sqrt{2\pi}} \frac{\sqrt{N}\left|\mu_0 - \overline{X}_N\right|}{(\sigma^2)^{3/2}} \end{aligned} \tag{3.29}$$

From the Law of the Iterated Logarithm, we have

$$\underset{N\to\infty}{Lim}\sup\left(\frac{\sqrt{N}\left|\mu_0-\overline{X}_N\right|}{\sigma^*\sqrt{2\log\log N}}\right)=1 \quad \text{a.e.} \quad P_{\theta_{10},\theta_2^*}, \tag{3.30}$$

$$\Rightarrow \qquad 0\le \underset{N\to\infty}{Lim}\{(\varepsilon_N)\sup_{\sigma^2,\left|\sigma^2-\sigma^{2*}\right|\le\varepsilon_N}\left|h_N(\underline{X}_N\mid\mu_0,\sigma^2)\right|\}$$

$$\le \frac{\underset{N\to\infty}{Lim}\,\varepsilon_N}{2\sqrt{\pi}}\left(\frac{\sqrt{\log\log N}}{(\sigma^{2*}-\varsigma)}\right),$$

$$\Rightarrow \qquad \underset{N\to\infty}{Lim}\frac{\varepsilon_N}{2\sqrt{\pi}}\left(\frac{\sqrt{\log\log N}}{(\sigma^{2*}-\varsigma)}\right)=0,\ P_{\theta_{10},\theta_2^*}\ \text{a.e.} \tag{3.31}$$

with $\sigma^{2*}>\varsigma>0$ .

Thus, both conditions of Theorem 2 are satisfied and therefore, the "average" p-value ( $\overline{p}$ ) for this example is consistent.

**<u>Example 3.2</u>**: Referring back to the separation hypothesis given in (2.3), let $\{X_i,\ i = 1,2,\ldots,K\}$ be independent with $\overline{X}_i$ ~N( $\mu_i,\sigma_i^2/n$ ), $i = 1,2,\ldots, K$, independent.

$$H_0\text{: } \sum_{i=1}^K l_i\mu_i\Big/\sqrt{\sum_{i=1}^K l_i^2\sigma_i^2}=n_\pi \quad \text{vs} \quad H_1\text{: } \sum_{i=1}^K l_i\mu_i\Big/\sqrt{\sum_{i=1}^K l_i^2\sigma_i^2}\ge n_\pi, \tag{3.32}$$

where $\{l_i\}$ are known constants. Letting $\underline{\theta}_1=(\mu_1,\mu_2,\ldots,\mu_K)$ and $\underline{\theta}_2=(\sigma_1^2,\sigma_2^2,\ldots,\sigma_K^2)$, $n_i/N=\lambda_i\in(0,1)$, we have here $\zeta(\underline{\theta})=\sum l_i\mu_i/\sqrt{\sum l_i^2\sigma_i^2}=n_\pi$. For known $\underline{\theta}_2$, a p-value may be based on

$$T_N(\underline{X}_N,\zeta,\underline{\theta}_2)=\left[\sum l_i\overline{X}_i-n_\pi\sqrt{\sum_{i=1}^K l_i^2\sigma_i^2}\right]\Big/\sqrt{\sum_{i=1}^K l_i^2\sigma_i^2/n_i},$$

which has a standard normal distribution under $H_0$.

(1) I present a Bayesian approach to testing these hypotheses using an average posterior p-value. Here, we use a Jeffreys-type noninformative prior, given by

$$p(\underline{\mu},\ \underline{\sigma}^2) = c(x)\frac{[\prod_{i=1}^{K} I_{(0,\infty)}(\sigma_i^2)]}{\prod_{i=1}^{K}\sigma_i^2} \quad .$$

The posterior distributions of $\underline{\mu}$ and $\underline{\sigma}^2$ are specified by

$$\mu_i \mid \sigma_i^2, \underline{x}_i \sim N(\bar{x}_i, \sigma_i^2 / n_i)$$

where $\{\mu_i \mid \sigma_i^2, \underline{x}_i\}_{i=1}^{K}$ are independent and

$$\sigma_i^2 \mid \underline{x}_i \underset{D}{\equiv} \frac{(n_i - 1)S_i^2}{U_i} \quad ,$$

$\{U_i \mid \underline{x}_i \sim \chi^2(n_i - 1)\}$ and $\{\sigma_i^2 \mid \underline{x}_i\}$ are independent with each other. So we have $\pi(\underline{\theta}_2 \mid \underline{x}) = \prod[\pi(\underline{\theta}_{2i} \mid \underline{x}_i)]$ and as in (3.27), $i = 1,2,\ldots,K$

$\theta_{2i} \mid \underline{x}_i = (n_i - 1)s_{n_i}^2 / U_i$, in distribution, $U_i \sim$ chi-square ($n_i$-1).

Consider the case $K \geq 2$, $n_i / N \to \lambda_i \in (0,1)$, $i = 1,2,\ldots,K$; and $H_0$: $\zeta(\underline{\theta}) = \zeta_0$ holds. Since $U_i = \sum_{j=1}^{m_i} Z_j^2$, $m_i = n_i - 1$, independent, $N(0,1)$ random variables,

$$Z_{m_i} = [U_i - m_i] / \sqrt{2m_i} \to N(0,1).$$

Hence, in distribution,

$$\theta_{2i} \mid \underline{x}_i = m_i s_{n_i}^2 / [Z_{m_i}\sqrt{2m_i} + m_i],$$

and

$$(\theta_{2i} - s_{n_i}^2) \mid \underline{x}_i = -(s_{n_i}^2 Z_{m_i}\sqrt{2/m_i}) / [Z_{m_i}\sqrt{2/m_i} + 1] \ . \tag{3.33}$$

Then, we have a.e., in distribution.

$$\sqrt{n_i}(\theta_{2i} - s_i^2) \mid \underline{x}_i \to N(0, 2\theta_{2i}^{*2}),\ \ i = 1,\ldots,K \ . \tag{3.34}$$

Note that $P(|s_i^2 - \theta_i^*| \geq \varepsilon_N \mid \underline{x}_N) = I(A_N)$, $i = 1,\ldots,K$, where

$$I(A_N) = 1 \text{ if } \left|s_i^2 - \theta_i^*\right| \geq \varepsilon_N \text{ and 0 otherwise.}$$

Take $\varepsilon_N = N^{q-.5}$. We have then,

$$\sum_{j=1}^{n_i}(X_{ij} - \overline{X}_i)^2 / n_i = \sum_{j=1}^{n_i}(X_{ij} - \mu_i)^2 / n_i - (\overline{X}_i - \mu_i)^2, \quad i = 1,\ldots,K .$$

Using the Law of the Iterated Logarithm and letting $Y_{ij} = (X_{ij} - \mu_i)^2$, $i = 1,\ldots,K$ and $j = 1,\ldots,n_i$

$$LimSup\left\{\varepsilon_N^{-1}\left|\sum_{j=1}^{n_i}(X_{ij} - \mu_i)^2 / n_i - \theta_{2i}^*\right|\right\} = LimSup\left\{N^{.5-q}\left|\sum_{j=1}^{n_i}(Y_{ij} - \theta_{2i}^*) / n_i\right|\right\}$$

$$= LimSup\{N^{.5-q}[|\sum_{j=1}^{n_i}(Y_{ij} - \theta_{2i}^*)| / n_i\sqrt{2n_i\theta_{2i}^* \log\log(n_i)}]\sqrt{2n_i\theta_{2i}^* \log\log(n_i)}\}$$

$$= LimSup\{[|\sum_{j=1}^{n_i}(Y_{ij} - \theta_{2i}^*)| / \sqrt{2n_i\theta_{2i}^* \log\log(n_i)}]\lambda_i^{-0.5+q} n_i^{-q}\sqrt{2\theta_{2i}^* \log\log(n_i)}\}$$

$$= 0, \text{ a.e. } P_{\theta^*}. \tag{3.35}$$

Likewise,

$$LimSup\{\varepsilon_N^{-1}(\overline{X}_i - \mu_i)^2\}$$

$$= LimSup\{n_i(\overline{X}_i - \mu_i)^2 / (N^q\sqrt{n_i\lambda_i})\} = 0, \text{ a.e. } P_{\theta^*}. \tag{3.36}$$

Hence,

$$LimSup\left\{\varepsilon_N^{-1}\left|\sum_{j=1}^{n_i}(X_{ij} - \overline{X}_i)^2 / n_i - \theta_{2i}^*\right|\right\} = 0, \text{ a.e. } P_{\theta^*}. \tag{3.37}$$

and

$$Lim\{P(\left|s_i^2 - \theta_{2i}^*\right| \geq \varepsilon_N \mid \underline{x}_i)\} = 0 \text{ a.e. } P_{\theta^*} \tag{3.38}$$

To verify condition (i) for the consistency of the average p-value, we then have :

$$P(\|\underline{\theta}_2 - \underline{\theta}_2^*\| \geq \varepsilon_N \mid \underline{x}_N)$$

$$\leq \sum_{i=1}^{K} P(|\theta_{2i} - \theta_{2i}^*| \geq \varepsilon_N / K \mid \underline{x}_N)$$

$$\leq \sum_{i=1}^{K} P(|\theta_{2i} - s_i^2| \geq \varepsilon_N / 2K \mid \underline{x}_N) + \sum_{i=1}^{K} P(|\theta_{2i}^* - s_i^2| \geq \varepsilon_N / 2K \mid \underline{x}_N)$$

$$\leq \sum_{i=1}^{K} P(\sqrt{n_i}(\theta_{2i} - s_i^2) \geq \sqrt{\lambda_i} N^q / 2K \mid \underline{x}_N) + \sum_{i=1}^{K} P(|\theta_{2i}^* - s_i^2| \geq N^{q-.5} / 2K \mid \underline{x}_N)$$

$$\to 0, \text{ a.e. }, \tag{3.39}$$

which verifies Condition (i) for the consistency of the average p-value.

(2) To verify (ii), letting $\underline{\theta} = ((\mu_{10}, \mu_{20}, ..., \mu_{K0}), \underline{\theta}_2)$ be any vector of parameters for which $H_0$ holds and $\Phi' = \phi$, the standard normal pdf, we obtain,

$$\overline{H}_N(\underline{X}_N \mid \zeta_0, \underline{\theta}_2) = \Phi\left(-\sum l_i(\overline{X}_i - \mu_{i0}) / \sqrt{\sum l_i^2 \sigma_i^2 / n_i}\right),$$

$$h_N^{(j)}(t \mid \underline{\theta}_{10}, \underline{\theta}_2) = \phi(-\sum l_i(\overline{X}_i - \mu_{i0}) / \sqrt{\sum l_i^2 \sigma_i^2 / n_i})[\frac{(1/2)\sum l_i(\overline{X}_i - \mu_{i0})}{n_j(\sum_i l_i^2 \sigma_i^2 / n_i)^{1.5}}] l_j^2$$

Let $\phi(-\sum l_i(\overline{X}_i - \mu_{i0}) / \sqrt{\sum l_i^2 \overline{\sigma}_i^2 / n_i}) = d \leq 1/\sqrt{2\pi}$ . Hence, using the Law of the Iterated Logarithm applied to $|\overline{X}_i - \mu_{i0}|$, $i = 1,2,\ldots,K$, we have that for positive constants $C_j$, $j = 1,\ldots, K$,

$$(\varepsilon_N) \sum_{j=1}^{K} \sup_{\underline{\theta}_2, \|\underline{\theta}_2 - \underline{\theta}_2^*\| \leq \varepsilon_N} \{h_N^{(j)}(\underline{X}_n \mid \underline{\theta}_{10}, \underline{\theta}_2)\}$$
$$= (\varepsilon_N) \sum_{j=1}^{K} \sup_{\underline{\theta}_2, \|\underline{\theta}_2 - \underline{\theta}_2^*\| \leq \varepsilon_N} \left| d[\frac{(1/2)\sum l_i(\overline{X}_i - \mu_{i0})}{n_j(\sum l_i^2 \sigma_i^2 / n_i)^{1.5}}] l_j^2 \right|$$

$$\le (N^{-0.5+q})\sum_{j=1}^{K}\frac{d\left|l_j^2\right|}{2n_j\left(\sum l_i^2\left(\sigma_i^{*2}-\varsigma\right)/n_i\right)^{1.5}}\ \sum_{i=1}^{K}\text{Lim sup}\left|[\frac{\sqrt{n_i}\left|l_i\right|\left|\bar{X}_i-\mu_{i0}\right|}{\sqrt{2\sigma_i^{*2}\log\log n_i}}]\right|\frac{\sqrt{2\sigma_i^{*2}\log\log n_i}}{\sqrt{n_i}}+o(1)$$

$$\le (N^{-0.5+q})\sum_{j=1}^{K}\frac{\sqrt{N}d\left|l_j^2\right|}{2\lambda_j\left(\sum l_i^2\left(\sigma_i^{*2}-\varsigma\right)/\lambda_i\right)^{1.5}}\ \sum_{i=1}^{K}\text{Lim sup}\left|[\frac{\sqrt{n_i}\left|l_i\right|\left|\bar{X}_i-\mu_{i0}\right|}{\sqrt{2\sigma_i^{*2}\log\log n_i}}]\right|\frac{\sqrt{2\sigma_i^{*2}\log\log N}}{\sqrt{\lambda_i N}}+o(1)$$

$$\le (N^{-0.5+q})\sum_{j=1}^{K}C_j\sqrt{\log\log N}+o(1)$$

$\to 0$, a.e. $P_{\underline{\theta}^*}$

with $n_j\sum l_i^2(\sigma_i^2/n_i) > \varsigma > 0$

Thus, both conditions of Theorem 2 are satisfied and therefore, the "average" p-value ( $\bar{p}$ ) for this example is consistent.

### 3.5 Posterior Predictive P-Values

The test statistic $T$ in equation (2.6) has two different levels of dependence on unknown (nuisance) parameters. The first level of dependence is on the nuisance parameter $\underline{\rho} = (1,\rho_2,...,\rho_K)$ in equation (2.6), which is equivalent to $\underline{\sigma}^2 = (\sigma_1^2,...,\sigma_K^2)$. The second level of dependence arises because the distribution of T depends on the unknown variances $\underline{\sigma}^2 = (\sigma_1^2,...,\sigma_K^2)$. In Chapter 2, one solution was to insert estimates for the nuisance parameters and take the maximum of p(( $\underline{\theta}_1,\underline{\theta}_2$ )) over the values of ( $\underline{\theta}_1,\underline{\theta}_2$ ) determined by the null hypothesis. As mentioned earlier, a traditional p-value does not exist in the cases I study because there is no useful pivotal.

Presenting a Bayesian view, Meng (1994) offered a solution to cases like this where the test variable depends on nuisance parameters by giving *an extended p-value*, called a *posterior predictive p-value* or *discrepancy p-value*, which is the tail area probability for a "discrepancy variable" under the joint posterior distribution of replicate data and the (nuisance) parameter, both conditional on the null hypothesis. The "discrepancy variable" is a "test statistic" dependent on unknown parameter(s). The posterior predictive p-value reduces the two levels of dependence. Following Meng (1994), given a null hypothesis $H_0 : \underline{\theta}_1 = \underline{\theta}_{10}$, the posterior predictive p-value is given by :

$$\begin{aligned} p_B &= \Pr\{D(\underline{x}^{rep}, \underline{\theta}_1, \underline{\theta}_2\} \geq D(\underline{x}, \underline{\theta}_1, \underline{\theta}_2) \mid \underline{x}, H_0\} \\ &= \Pr\{D(\underline{x}^{rep}, \underline{\theta}_{10}, \underline{\theta}_2\} \geq D(\underline{x}, \underline{\theta}_{10}, \underline{\theta}_2) \mid \underline{x}\} \end{aligned} \tag{3.40}$$

where $D(\underline{x}, \underline{\theta}_1, \underline{\theta}_2)$ is a discrepancy variable, $\underline{x}^{rep}$ denotes a replication of $\underline{x}$, a "future observation". The probability in (3.40) is taken over the joint posterior distribution of $(\underline{x}^{rep}, \underline{\theta}_{10}, \underline{\theta}_2)$ given $H_0$. Specifically,

$$\begin{aligned} f(\underline{x}^{rep}, \underline{\theta}_1, \underline{\theta}_2 \mid \underline{x}, H_0) &= f(\underline{x}^{rep} \mid \underline{\theta}_1, \underline{\theta}_2)\pi_0(\underline{\theta}_2 \mid \underline{x}), \quad \underline{\theta}_1 = \underline{\theta}_{10} \ , \\ &= f(\underline{x}^{rep} \mid \underline{\theta}_{10}, \underline{\theta}_2)\pi_0(\underline{\theta}_2 \mid \underline{x}) \end{aligned} \tag{3.41}$$

where $\pi_0(\underline{\theta}_2 \mid \underline{x})$ is the posterior density (probability) of $\underline{\theta}_2$ under $H_0$. Meng also gives an alternative interpretation of $p_B$ by taking the posterior mean of $p(\underline{\theta}_2)$ over the posterior distribution of $\underline{\theta}_2$ under $H_0$. That is,

$$p_B = E(p(\underline{\theta}_2) \mid \underline{x}, H_0) \ , \tag{3.42}$$

where $p(\underline{\theta}_2) = \Pr\{D(\underline{X}, \underline{\theta}_1, \underline{\theta}_2\} \geq D(\underline{x}, \underline{\theta}_1, \underline{\theta}_2) \mid \underline{\theta}_1 = \underline{\theta}_{10}, \underline{\theta}_2\}$. This probability is obtained from the frequentist setting, using the sampling density $f(X \mid \underline{\theta}_{10}, \underline{\theta}_2)$.

Choosing discrepancy variables can be difficult. Meng(1994) suggested two discrepancy variables, called a *conditional likelihood ratio* (CLR) and a *generalized likelihood ratio* (GLR), assuming that the density $f(\cdot)$ is jointly continuous in its arguments, is given by

$$D^{C}(\underline{x},\underline{\theta}_2)=\frac{\sup_{\underline{\theta}_1\notin\Theta_0} f(\underline{x}\,|\,\underline{\theta}_1,\underline{\theta}_2)}{\sup_{\underline{\theta}_1\in\Theta_0} f(\underline{x}\,|\,\underline{\theta}_1,\underline{\theta}_2)} \tag{3.43}$$

$$D^{G}(\underline{x})=\frac{\sup_{\underline{\theta}_1\notin\Theta_0}\sup_{\underline{\theta}_2} f(\underline{x}\,|\,\underline{\theta}_1,\underline{\theta}_2)}{\sup_{\underline{\theta}_1\in\Theta_0}\sup_{\underline{\theta}_2} f(\underline{x}\,|\,\underline{\theta}_1,\underline{\theta}_2)} \tag{3.44}$$

Meng (1994) used two classical examples, including the Behrens-Fisher problem, to illustrate the posterior predictive p-value. A posterior predictive p-value need not have a uniform distribution under $H_0$. But, Meng (1994) shows that if the replication is defined by nuisance parameters and new data generated, then the Type I frequentist error of a nominal $\alpha$-level posterior predictive test is often close to but less than $\alpha$ and will never exceed $2\alpha$.

Here, I verify that the posterior predictive p-value with the conditional likelihood ratio (CLR), the discrepancy variable, is exactly equivalent to the posterior p-value with the test statistics in (3.19) and is then also equivalent to $P(H_0\,|\,Data)$.

To prove that $p_B$ is equivalent to $P(H_0\,|\,Data)=\overline{p}_{post}$, we need to obtain the relationship between CLR and $Z(\underline{\sigma}^2)$ as defined in (3.19). In fact, it is easy to check that the CLR is a monotone function of $Z(\underline{\sigma}^2)$.

**<u>Verification 3.5.1</u>**: The log likelihood function is given by

$$\log L(\underline{\mu},\underline{\sigma}^2)=-\sum_{i=1}^{K}\frac{n_i}{2}\log(2\pi\sigma_i^2)-\sum_{i=1}^{K}\sum_{j=1}^{n_i}(x_{ij}-\mu_i)^2/2\sigma_i^2$$

Suppose $\underline{\theta}_2=\underline{\sigma}^2$ is fixed. Then, it is easy to obtain the maximum likelihood estimators (MLE's) under the null hypothesis and alternative hypothesis.

For $\underline{\theta}_1\notin\Theta_0$, the MLE is : $\tilde{\mu}_i=\overline{x}_i$ (sample mean)

For $\underline{\theta}_1\in\Theta_0$, the MLE is : $\hat{\mu}_i=\overline{x}_i+\dfrac{l_i\sigma_i^2[n_\pi\sqrt{\sum_{i=1}^{K}l_i^2\sigma_i^2}-\sum_{i=1}^{K}l_i\overline{x}_i]}{n_i\sum_{i=1}^{K}\dfrac{l_i^2\sigma_i^2}{n_i}}$

Then, we can obtain the conditional likelihood ratio (CLR):

$$D^{C}(\underline{x},\underline{\theta}_2)=\frac{\sup_{\underline{\theta}_1\notin\Theta_0} f(\underline{x}\mid\underline{\theta}_1,\underline{\theta}_2)}{\sup_{\underline{\theta}_1\in\Theta_0} f(\underline{x}\mid\underline{\theta}_1,\underline{\theta}_2)}$$

$$=\exp[\sum_{i=1}^{K}(\sum_{j=1}^{n_i}(\frac{(x_{ij}-\hat{\mu}_i)^2-(x_{ij}-\tilde{\mu}_i)^2}{2\sigma_i^2}))]$$

$$=(\frac{1}{2})\exp\left[\frac{[\sum l_i\bar{x}_i-n_\pi\sqrt{\sum l_i^2\sigma_i^2}\,]^2}{\sum l_i^2\sigma_i^2/n_i}\right]$$

$$=(\frac{1}{2})\exp\left([Z(\underline{\sigma}^2)]^2\right).$$

Thus, $\quad 2\ln[D^{C}(\underline{x},\underline{\theta}_2)]\sim\chi^2(1)$ (a chi-square distribution with *df* = 1)

$\Rightarrow p_B$ is equivalent to $P(H_0\mid Data)=\overline{p}_{post}$,

which completes the verification.

# Chapter 4 Nonparametric Tests

The Wilcoxon-Rank sum test, also called the Mann-Whitney test (Wilcoxon 1945; Mann and Whitney 1947), provides an exact size $\alpha$ test for the equality of two continuous distributions, denoted $F_1$ and $F_2$, based on independent random samples. It is the locally most powerful rank test for detecting a shift in the logistic distribution and has good power for many other shift models, without the need to assume a particular distributional form. The Mann-Whitney form of the test indicates that it is based on estimating $\pi_{xy} \equiv P(X > Y)$, $X \sim F_1$ and $Y \sim F_2$, and its asymptotic power function increases as $\pi_{xy}$ moves away from 1/2. However, the test's null distribution is obtained under the assumption that $F_1 \equiv F_2$ and it can perform poorly as a test of equal locations when $F_1$ and $F_2$ are not just translates of one another. The rank Welch test (Zimmerman and Zumbo 1993) provides an approximate size $\alpha$ test of the stochastic equality and inequality (4.1) using an approximating Student-t distribution, but it exhibits some $\alpha$ inflation in certain cases as given in Delaney and Vargha (2002). To deal with this case, Reiczigel, Zakaria's and Ro'zsa (2005) developed a new test, called the *Bootstrap Rank Welch test* (*BRW*), to test for stochastic symmetry without assuming that the distributions have the same shape. Their hypotheses are given by

$$H_0 : P(X < Y) = P(X > Y), \quad H_a : P(X < Y) \neq P(X > Y), \tag{4.1}$$

for two-sided tests, and

$$H_{a1} : P(X < Y) > P(X > Y) \ \text{ or } \ H_{a2} : P(X < Y) < P(X > Y), \tag{4.2}$$

for a one-sided test.

As described below, I will extend their hypotheses and construct tests to deal with three or more distributions. Also, Teprstra(1952) and Jonckheere(1954) proposed a nonparametric test for ordered alternatives among two or more distributions based on the sum of pairwise Mann-Whitney statistics. In future work I plan to extend their null hypothesis of equality among the distributions to encompass a degree of ordered separation.

A generalization of (4.1) relevant to my goal of assessing separation is $IS_{AV}(\underline{F})$ given in (1.8) for some user input $\pi > ½$,

$$H_0 : 2\sum_{i<j} \max\{\pi_{ij}, \pi_{ji}\}/(K(K-1)) \leq \pi \,, \quad H_a : 2\sum_{i<j} \max\{\pi_{ij}, \pi_{ji}\}/(K(K-1)) > \pi \,. \tag{4.3}$$

I also propose the related hypotheses based on $IS_{MAX}(\underline{F})$ given in (1.7),

$$H_0 : \max\{\pi_{ij}\} \leq \pi \,, \quad H_a : \max\{\pi_{ij}\} > \pi \,. \tag{4.4}$$

Another well known extension of (4.1) is given by what are called *slippage tests*, which will be discussed further in section 4.4.

### 4.1 A Nonparametric Test Statistic

Developing exact tests for (4.3)-(4.4) is not possible in general since the null hypotheses are composite and do not require that the distributions be identical. Instead, I will develop tests based on the bootstrap and study their properties in terms of size, power and robustness. For example, a test for (4.4) could be based on the statistic

$$\hat{\tau} = \max_{i \neq j, k=1,\ldots,n_i, l=1,\ldots,n_j} \left[ \frac{\#(x_{ik} > x_{jl})}{n_i n_j} \right] \,. \tag{4.5}$$

To construct bootstrap tests, we need to take resamples from an estimate of $F_0$, the distribution of the data under the null hypothesis. For composite null hypotheses, where $H_0$ does not fully specify $F_0$, Efron and Tibshirani (1993) propose the following guidelines:

1. Use a test statistic which is approximately pivotal so that its distribution changes little over the conditions determined by the null hypothesis.
2. Condition on a sufficient statistic for the unknown parameters.
3. Estimate $F_0$ by a CDF $\hat{F}_0$ which satisfies $H_0$ and resample from it.

Here, I will focus on (3). In our hypothesis tests, there are many parameters of interest and many nuisance parameters, so that constraining $F$ so that the null hypothesis holds will be very challenging.

## 4.2 Nonparametric Bootstrap Tests

The proposed method applies the nonparametric bootstrap principle to testing (4.4) based on the new test statistic $\hat{\tau}$ given in (4.5). First we need to transform the samples $\{\underline{x}_1, \underline{x}_2, \ldots \underline{x}_K\}$ into $\{\underline{x}_1', \underline{x}_2', \ldots \underline{x}_K'\}$ to satisfy the null hypothesis, that is, to stochastic equality. The null distribution of $\hat{\tau}$ is estimated by resampling from the distributions $(\underline{x}_1', \underline{x}_2', \ldots \underline{x}_K'\}$ and calculating the test statistic for each resample group ($\hat{\tau}^*_{(b)}$, b = 1,2,…,$B$). Based on this simulated null distribution of $\hat{\tau}$, a P-value can be defined as:

$$p_1 = \frac{1}{B}\sum 1(\hat{\tau} \ge \hat{\tau}^*_{(b)}) \qquad \text{for an upper-tail test} \tag{4.6}$$

$$p_2 = \frac{1}{B}\sum 1(\hat{\tau} \le \hat{\tau}^*_{(b)}) \qquad \text{for a lower-tail test} \tag{4.7}$$

$$p = 2\min\{p_1, p_2\} \qquad \text{for a two-tailed test} \tag{4.8}$$

The challenging part here is to figure out how to transform the samples into a new data set to satisfy the null hypothesis, a problem I will work on.

### 4.2.1 Symmetric Distributions

If the distributions $\{F_i\}$ are symmetric and $\pi = 1/2$, the hypotheses in inequalities (4.3) and (4.4) may be viewed as a test for the equality of means,

$$H_0 : \mu_1 = \mu_2 = \ldots = \mu_K,$$

$$H_a : \text{At least two means are different.} \tag{4.9}$$

Then we could just use the usual shift transformation. First, compute an overall mean for the pooled data, denoted by $\bar{x}$. Second, use the equation below to complete the transformation.

$$x_{ij}' = x_{ij} - \bar{x}_i + \bar{x} \quad , \; j = 1,2,...,n_i \, , \; i = 1,2,...,K \; , \tag{4.10}$$

where $\bar{x}_i = \sum_{j=1}^{n_i} x_{ij} / n_i$ is the sample mean of the *i* th sample.

### 4.2.2 Skewed Distributions

- **Shift Models**

Assume that we have two or more independent random variables with the same shape and $\pi = 1/2$. Then our hypothesis test (4.3) and (4.4) becomes a test of equality of distributions.

We might follow the shift transformation (4.11), use the sample median instead of sample mean to minimize the effect of outliers,

$$x_{ij}' = x_{ij} - Med_i + Med \, , \quad j = 1,2,...,n_i \, , \; i = 1,2,...,K \; , \tag{4.11}$$

where $Med_i$ denotes the sample median of the *i th* variable, and $Med$ denotes the overall sample median. A simulation study is needed in this case.

- **General Models**

If we know nothing about the distributions of those random variables, then figuring out how to transform the samples into a new data set to satisfy the null hypothesis is a difficult problem. In the present study, there are three potentially useful transformations to try. The first one is the shift transformation (4.12), a little different from (4.10). The second transformation is called a stretch transformation (4.13). The last one is a power transformation (4.14).

$$x_{ij}' = x_{ij} + a \, , \quad j = 1,2,...,n_i \, , \; i = 2,...,K \; , \tag{4.12}$$

where $a$ can be obtained as the median of the values ( $x_{ij} - x_{1m}$ ).

$$x_{ij}' = c(x_{ij} - w) + w, \quad j = 1,2,...,n_i,\ i = 2,...,K, \tag{4.13}$$

where $w \geq 0$ and $c$ can be obtained as the median of the values ( $\dfrac{x_{ij} - w}{x_{1m} - w}$ ).

$$x_{ij}' = x_{ij}^d \quad j = 1,2,...,n_i,\ i = 2,...,K, \tag{4.14}$$

where $d$ can be obtained as the median of values { $\dfrac{\log(x_{ij})}{\log(x_{1m})}$ }.

For those three transformations, we let $x_{1m}' = x_{1m}$, $m = 1,2,...,n_1$, and make some transformations for the other samples to satisfy the null hypothesis. Note that the last two transformations change the ratio of the variances. We need to do a simulation study in the future to see if this works.

### 4.2.3 A Nonparametric Bootstrap CI for ISP

Since transforming the samples { $\underline{x}_1, \underline{x}_2, ... \underline{x}_K$ } into { $\underline{x}_1', \underline{x}_2', ... \underline{x}_K'$ } to satisfy the null hypothesis is complicated, especially to our composite hypothesis. I will use a nonparametric bootstrap to construct a one-sided lower confidence set *CI* for $IS_{AV}(\underline{F})$ and $IS_{MAX}(\underline{F})$, and conclude that the alternative hypothesis is correct if *CI* does not contain $\Psi$. The nonparametric bootstrap CI for $IS_{MAX}(\underline{F})$ is carried out as follows

(1) Resample independently and separately from the data $\{x_{ij}\}_{j=1}^{n_i}$ $i = 1,...,K$, and compute $\hat{IS}_{MAX}(\underline{F}) = \hat{\tau}$.

(2) Independently repeat *B* times, resulting in $\{\hat{\tau}_i^*\}_{i=1}^{B}$.

(3) Use the percentile method or prepivoting method developed by (Beran, 1987) to construct a Lower CI for $IS_{MAX}(\underline{F})$: $CI = [\hat{L}(Data), \infty]$. Reject $H_0$ if $\pi < \hat{L}(Data)$.

Simulation results for this procedure are given in Chapter 5. In future research, I will use the *BCa* (Bias Corrected and Accelerated) bootstrap to construct a Lower CI for $IS_{AV}(\underline{F})$ and $IS_{MAX}(\underline{F})$.

### 4.3 The Quantile Test Statistic

I examined another test statistic for the hypothesis tests (4.2)-(4.4) called the *quantile test statistic*, which is given by:

$$H(p) = \frac{\max_{i \neq j} \left| F_i^{-1}(p) - F_j^{-1}(p) \right|}{\frac{1}{K} \sum_{i=1}^{K} \left| F_i^{-1}(3/4) - F_j^{-1}(1/4) \right|} \tag{4.15}$$

where $F_i^{-1}(p)$ is the inverse CDF for ith treatment, defined by:

$$F_i^{-1}(p) = \inf\{ y : F_i(y) \geq p \} \quad \text{for } 0 \leq p \leq 1. \tag{4.16}$$

A preliminary simulation was conducted and the results did not show a general pattern in a plot of the test statistic $H(p)$ verses $p$. Therefore, the quantile test statistic needs to be revised or adjusted in future research.

### 4.4 Slippage Tests

Slippage tests were considered as an outlier distribution detection by Mosteller (1948), Paulson (1952), Kudo (1956), Doornbos and Prins (1956), and others for location slippage, and by Cochran (1941) for variance slippage. In general, suppose we want to compare *K* distributions to find out if all these distributions are identical, or, if not, which one has “slipped” away from the others, which are identical. Actually, this is a more restrictive test for both null and alternative hypotheses, since it only considers one ‘extreme’ distribution. R. Doornbos published a book called “Slippage tests” in 1966 to describe the slippage tests for one, or more than one outlier under several families of distributions.

In my hypothesis test (4.1), another extension is given by *slippage tests*, whose *right-sided* hypotheses are given by:

$H_0: \; P(X_i < X_j) = P(X_i > X_j) = \pi_{ij} = 1/2 \quad \text{for } i \neq j$

$H_{a1}: P(X_i > X_j) = \pi_{ij} > \dfrac{1}{2} \qquad (\mathit{for} \;\; j \neq i = m, \;\; \mathit{and} \;\; j = 1,\ldots,m-1, m+1,\ldots,K; m \text{ unknown})$

$\mathit{and} \;\; P(X_i > X_j) = \; P(X_i < X_j) = \pi_{\mathrm{ij}} = \dfrac{1}{2} \quad (\text{for } j \neq i, \;\; \mathit{and} \;\; i, j = 1,\ldots,m-1, m+1,\ldots,K)$

(4.17)

Left-sided slippage hypotheses are given by

$H_{a2}: P(X_i > X_j) = \pi_{ij} < \dfrac{1}{2} \qquad (\mathit{for} \;\; j \neq i = m, \;\; \mathit{and} \;\; j = 1,\ldots,m-1, m+1,\ldots,K)$

$\mathit{and} \;\; P(X_i > X_j) = P(X_i < X_j) = \pi_{ij} = \dfrac{1}{2} \;\; (\text{for } j \neq i, \;\; \mathit{and} \; i, \; j = 1,\ldots,m-1, m+1,\ldots,K)$

(4.18)

If we assume that those random variables $\{X_i\}$ have the same shape, then the hypothesis tests in equation (5.5) and (5.6) become.

$$\begin{aligned} &H_0: F_1 = F_2 = \ldots = F_K, \\ &H_{a1}: P(X_i > X_j) = \pi_{ij} > \frac{1}{2} \qquad (\mathit{for} \;\; i \neq j) \\ &\mathit{and} \;\; X_j \; (j = 1,\ldots,i-1, i+1,\ldots,K) \;\; \mathit{follow \;\; the \;\; same \;\; distribution} \end{aligned} \tag{4.19}$$

for one unknown value of *i* (right-slippage test), and

$$\begin{aligned} &H_{a2}: P(X_i > X_j) = \pi_{ij} < \frac{1}{2} \qquad (\mathit{for} \;\; i \neq j) \\ &\mathit{and} \;\; X_j \; (j = 1,\ldots,i-1, i+1,\ldots,K) \;\; \mathit{follow \;\; the \;\; same \;\; distribution} \end{aligned} \tag{4.20}$$

for one unknown value of *i* (left-slippage test).

The test statistics $\hat{\tau}$ in (4.5) and $H(p)$ in (4.15) can also can be used to test for the hypotheses in (4.17) – (4.20).

# Chapter 5 Simulation Results and Discussion

Before carrying out a full simulation study, I did a small-scale simulation to compare the bias of the likelihood ratio test (LRT) of the hypotheses given in (2.3), assuming normality, using Method 1 and Method 2, as explained in Section 2.5, Let $\pi_0$ denote the maximum value of $\pi$ under $H_0$. Estimated power functions for selected values of $\pi \le \pi_0$ ( $H_0$ holds) are reported in Appendix B, Table B.1, for the case $\pi = \pi_0$ and Table B.2 for $\pi < \pi_0$ , where the LRT p-value is calibrated using a chi-square distribution with df=1 and PBL p-values are calibrated using a bootstrap. I also present QQ plots of the p-values (*LRT* and *PBL*) vs a uniform (0, 1) distribution for the two methods in Appendix B, Figure B.1.

From these tables and graphs we see that power functions using Method 2 are close to and mostly less than the nominal $\alpha$ =0.05 value for all cases. On the other hand, the entries for Method 1 are much larger than nominal when $\pi < \pi_0$. These conclusions are supported by the QQ plots in Appendix B, Figure B.1 which show sharp departures from linearity unless $\pi = \pi_0$.

In addition, to investigate the distribution of the test statistics $\lambda$, the logarithm of the likelihood ratio test statistic, Figure B.2 in Appendix B presents histograms of simulated , independent copies of $\lambda$ under several conditions when $H_0$ holds. Sample means and variances of these histograms indicate sometimes significant departures from the values of one and two, respectively, which would be the case if these were samples from a chi-square distribution with one degree of freedom. Consequently, the chi-square distribution with one degree of freedom should not be used to calibrate the LRT for this class of hypotheses and accordingly my full-scale simulation only uses the Method 2 with LRT statistic calibrated using a bootstrap, designated *PBL*.

Furthermore, I investigated the affect of different choices of the values $\{\pi_j\}$ used in Method 2 on the behavior of the PBL. The results in Table 5.1 and the QQ plots in Appendix B, Figure B.3, indicate that there is not a large difference due to changing the gap value, $|\pi_j - \pi_{j+1}|$ from 0.01 to 0.05. In my full simulations described below, I will use three different gaps.

Table 5.1 Estimated Type I Error Probabilities Comparison of *PBL* P-Values for $IS_{LIN}(\underline{F})$
$K$=3, $\underline{l}^T = (-1/2, -1/2, 1)$, (b) $(\mu_2, \mu_3) = (0.5, 2)$, $\alpha$ =0.05, $\pi = \pi_0 = 0.75$ (*PBL*),
Iterations = 1000

| *Gap* | Small sample sizes | | Medium sample sizes | | large sample sizes | |
|---|---|---|---|---|---|---|
| | $\alpha = 0.05$ | $\alpha = 0.10$ | $\alpha = 0.05$ | $\alpha = 0.10$ | $\alpha = 0.05$ | $\alpha = 0.10$ |
| 0.01 | 0.05 | 0.09 | 0.035 | 0.065 | 0.04 | 0.09 |
| 0.025 | 0.05 | 0.09 | 0.035 | 0.065 | 0.04 | 0.09 |
| 0.05 | 0.05 | 0.09 | 0.035 | 0.065 | 0.04 | 0.09 |

### 5.1 Simulation Study for $IS_{LIN}$ (F) Assuming Normality

In this sub-section, I will focus on the test for the hypotheses given in (2.3). Simulations here were used to check and compare the Type I error rates and power curves of five p-values for $IS_{LN}$ (F) under normality. Recall that

$$IS_{LIN}(\underline{F}) = \sum_{i=1}^{K} l_i \mu_i \Big/ \sqrt{\sum_{i=1}^{K} l_i^2 \sigma_i^2} \quad .$$

I consider the fixed-effects model

$$X_{ij} = \mu_i + \varepsilon_{ij} ,$$

where the independent error terms $\varepsilon_{ij} \sim N(0, \sigma_i^2), i = 1,2,...,K; j = 1,...,n_i$ , $N = \sum n_i$ , resulting in data $\{x_{ij}, j = 1,2,\ldots,n_i\}$ .

**<u>Parameter Settings</u>:**

$K = 3, 5, 7$

$\underset{\sim}{l}^T = (-1/(K-1), -1/(K-1), \ldots, 1)$, $(-1/(J_1), -1/(J_1), \ldots, 1/(J_2), 1/(J_2))$

where $\sum_{\#(J_1)} 1/J_1 = \sum_{\#(J_2)} 1/J_2$, and $J_1, J_2 < K$ are specified below,

$\alpha = 0.05, 0.10$

$\mu_1 = 0$, $\sigma_1^2 = 1$ (without loss of generality)

$\# = 200$ iterations (data sets)

sample sizes : $n$ = 10(small), 30(medium), 100 (large)

Average p-value: generate 1000 independent variances from the chi-square distribution.

Bootstrap procedure: generate 99 bootstrap samples.

***Case 1***

- **K=3:** $\underset{\sim}{l}^T = (-1/2, -1/2, 1)$

$\pi = 0.55, 0.65, 0.75, 0.85$

(a) Equal Variance: $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = 1$

$(\mu_2) =$ **(0), (0.5), (1)**--------3 parameter settings

$\pi = 0.75, 0.8, 0.85, 0.9$ for $(\mu_2) =$ **(1)**

Obtain $\mu_3$ so that $\Phi[\sum_{i=1}^{K} l_i \mu_i / \sqrt{\sum_{i=1}^{K} l_i^2 \sigma_i^2}] = \pi$.

(b) $\sigma_2^2 = \mu_2 + 1$, so that the variance increases with the mean.

$(\mu_2, \mu_3) =$ **(0.5, 1), (0.5, 2) , (0.5, 5)**------3 parameter settings

$\pi = 0.55, 0.65, 0.75, 0.8$ for $(\mu_2, \mu_3) =$ **(0.5, 1)**

Obtain $\sigma_3^2$ so that $\Phi[\sum_{i=1}^{K} l_i \mu_i / \sqrt{\sum_{i=1}^{K} l_i^2 \sigma_i^2}] = \pi$.

***Case 2***

- **K=5:** $\underset{\sim}{l}^T = (-1/3, -1/3, -1/3, 1/2, 1/2)$

$\pi = 0.75, 0.8, 0.85, 0.9$

(a) Equal Variance: $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2 = \sigma_5^2 = 1$

$(\mu_2, \mu_3, \mu_4) =$ **(0, 0, 1), (0.5, 0.5, 1)** --------2 parameter settings

Obtain $\mu_5$ so that $\Phi[\sum_{i=1}^{K} l_i\mu_i / \sqrt{\sum_{i=1}^{K} l_i^2\sigma_i^2}] = \pi$ .

(b) $\sigma_i^2 = \mu_i + 1$ for $i$ =2,3,4

$(\mu_2, \mu_3, \mu_4, \mu_5) =$ **(0, 0, 2, 2), (0.5, 1, 3, 3.5)** ------2 parameter settings

Obtain $\sigma_5^2$ so that $\Phi[\sum_{i=1}^{K} l_i\mu_i / \sqrt{\sum_{i=1}^{K} l_i^2\sigma_i^2}] = \pi$ .

***Case 3***

- **K=7:** $\underline{l}^T = (-1/6, -1/6, -1/6, -1/6, -1/6, -1/6, 1)$

(a) Equal Variance: $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2 = \sigma_5^2 = \sigma_6^2 = \sigma_7^2 = 1$

$\pi$ = 0.75, 0.8, 0.85, 0.9

$(\mu_2, \mu_3, \mu_4, \mu_5, \mu_6) =$ **(0,0,0,0,0), (0.5, 0.5, 0.5, 1,1), (1, 1, 1, 1, 1)**

--------3 parameter settings

Obtain $\mu_7$ so that $\Phi[\sum_{i=1}^{K} l_i\mu_i / \sqrt{\sum_{i=1}^{K} l_i^2\sigma_i^2}] = \pi$ .

(b) $\sigma_i^2 = \mu_i + 1$ for $i$ =2,3,4,5,6

$\pi$ = 0.55, 0.65, 0.75, 0.85

$(\mu_2, \mu_3, \mu_4, \mu_5, \mu_6, \mu_7) =$ **(0, 0, 0, 0, 0, 1), (0, 0.5, 0.5, 0.5, 0.5, 2) ,**

**(0.5, 1, 1, 1, 1.5, 3)**------3 parameter settings

Obtain $\sigma_7^2$ so that $\Phi[\sum_{i=1}^{K} l_i\mu_i / \sqrt{\sum_{i=1}^{K} l_i^2\sigma_i^2}] = \pi$ .

For each of the parameter combinations, I compare the following test statistics for both size and the power. In each case, the null hypothesis is rejected if the estimated p-value is at most the nominal type I error rate, $\alpha$ .

Test Statistic: ***Plug***: Plug-in p-value for $T = \dfrac{\sum l_i \overline{X}_i}{\sqrt{\sum l_i^2 S_i^2 / n_i}}$,

***A_T:*** Average p-value for $T(\underline{\sigma}^2)$

***A_Z:*** Average p-value for $Z(\underline{\sigma}^2)$ = Posterior Predictive p-value

***PBp:*** Parametric Boostrap Test p-value for plug-in $T = \dfrac{\sum l_i \overline{X}_i}{\sqrt{\sum l_i^2 S_i^2 / n_i}}$

***PBL:*** Parametric Boostrap Test p-value for LRT statistic $\lambda$

The Type I error rates and the power comparisons for $\alpha = 0.05$ and 0.10 are estimated using Monte Carlo simulation. As mentioned above, I will use Method 2 throughout my simulation study to generate *PBL* p-values under the null hypothesis. I set the gaps between $\pi_j$ and $\pi_{j+1}$ equal to three cases, 0.01, 0.025, 0.05.

### 5.1.1 Comparison of Type I Error Rates

In (2.3), $H_0$ is composite for $IS_{LIN}(\underline{F})$, which makes it difficult to check type I error rates and powers . Table 5.2 reports representative cases of estimated type I error rates, $\hat{\alpha}$, for these five tests with small, medium, and large sample sizes and nominal type I error rate $\alpha = 0.05$ when $\pi = \pi_0$. The complete simulation results are summarized in Appendix C, Table C.1. Cells in the tables where $\alpha = 0.05$ does not lie in the approximate .95 confidence interval $\hat{\alpha} \pm 1.96\sqrt{\hat{\alpha}(1-\hat{\alpha})/200}$ are highlighted. The gray color indicates that the corresponding entry is smaller than the lower bound of the approximate .95 confidence interval. The pink color indicates that the corresponding entry is greater than the upper bound of the CI above. The results for $\alpha = 0.10$ are very similar and are not reported. Further, some QQ plots for these five p-values are given in Appendix C, Figure C.1. These tables and plots lead to the following summary statements.

*(1)* In general, when the sample sizes are large, the Type I error rates for all five tests are very similar and close to the nominal level $\alpha$ .

*(2)* For small samples, the Type I error rates for *Plug* may exceed the nominal $\alpha$ in some cases, but appear never to exceed $2\alpha$ . But, for most of cases, even for small samples, the estimated Type I error rates for *Plug* p-value tests appear to be more stable and closer to the nominal level $\alpha$ than the error rates for the others.

*(3)* The Average p-value tests for *T (A_T)* and for *Z (A_Z)* have similar estimated Type I error rates and these two p-values seem to be a little conservative for small samples.

*(4)* The QQ plots in Appendix C, Figure C.1, appear to be equiangular lines through the origin, especially for the large samples.

*(5)* Furthermore, in order to check whether the estimated type I error probabilities for those five tests are less than the nominal $\alpha$ over the parameter space determined by the composite null hypothesis, I generated data sets under $\pi < \pi_0$ . The corresponding results are exhibited in Appendix C, Table C.2 for $\alpha$ =0.05 and in Appendix C, Figure C.2 using QQ plots. These simulation results indicate that the estimated levels of those five tests are less than the nominal $\alpha = 0.05$. Specifically, for some small samples even some medium-size samples, and the large values of $\pi$ , the *PBp* and *PBL* p-values have a very conservative Type I error rate. Sometimes, the estimated Type I error rate = 0.

Although the *PBp* and *PBL* test compare well with the tests based on the *plug in* and *average p-values*, they are complex, time-consuming procedures.

Table 5.2 Estimated Type I Error Probabilities Using P-Values for $IS_{LIN}(\underline{F})$, $\pi = \pi_0$.

Table 5.2.1 $K$=3, (a) $(\mu_2) = (0.5)$, $\alpha$ =0.05, $\underline{l}^T = (-1/2, -1/2, 1)$

| $\pi$ | small sample sizes | | | | | medium sample sizes | | | | | large sample sizes | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *Plug* | *A_T* | *A_Z* | *PBp* | *PBL* | *Plug* | *A_T* | *A_Z* | *PBp* | *PBL* | *Plug* | *A_T* | *A_Z* | *PBp* | *PBL* |
| *0.55* | 0.06 | 0.05 | 0.05 | 0.065 | 0.065 | 0.065 | 0.06 | 0.065 | 0.06 | 0.06 | 0.045 | 0.045 | 0.045 | 0.05 | 0.05 |
| *0.65* | 0.04 | 0.04 | 0.04 | 0.045 | 0.05 | 0.055 | 0.05 | 0.05 | 0.055 | 0.06 | 0.05 | 0.05 | 0.05 | 0.05 | 0.06 |
| *0.75* | 0.055 | 0.03 | 0.03 | 0.055 | 0.03 | 0.065 | 0.05 | 0.05 | 0.06 | 0.04 | 0.04 | 0.04 | 0.04 | 0.045 | 0.035 |
| *0.85* | 0.06 | 0.05 | 0.05 | 0.036 | 0.025 | 0.055 | 0.055 | 0.055 | 0.055 | 0.05 | 0.055 | 0.05 | 0.05 | 0.06 | 0.045 |

Table 5.2.2 $K$=3, (b) $(\mu_2, \mu_3) = (0.5, 2)$, $\alpha$ =0.05, $\underline{l}^T = (-1/2, -1/2, 1)$

| $\pi$ | small sample sizes | | | | | Medium sample sizes | | | | | large sample sizes | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *Plug* | *A_T* | *A_Z* | *PBp* | *PBL* | *Plug* | *A_T* | *A_Z* | *PBp* | *PBL* | *Plug* | *A_T* | *A_Z* | *PBp* | *PBL* |
| *0.55* | 0.04 | 0.045 | 0.045 | 0.04 | 0.05 | 0.035 | 0.035 | 0.04 | 0.045 | 0.045 | 0.05 | 0.05 | 0.05 | 0.045 | 0.045 |
| *0.65* | 0.035 | 0.035 | 0.035 | 0.035 | 0.055 | 0.06 | 0.055 | 0.06 | 0.06 | 0.06 | 0.045 | 0.045 | 0.045 | 0.045 | 0.03 |
| *0.75* | 0.06 | 0.055 | 0.055 | 0.06 | 0.04 | 0.05 | 0.05 | 0.05 | 0.055 | 0.06 | 0.05 | 0.05 | 0.05 | 0.055 | 0.04 |
| *0.85* | 0.045 | 0.035 | 0.035 | 0.016 | 0.021 | 0.065 | 0.055 | 0.055 | 0.036 | 0.026 | 0.03 | 0.03 | 0.03 | 0.03 | 0.055 |

### 5.1.2 Power Comparisons

To compare powers $\kappa(\underline{\theta})$ at specified alternatives, I present estimated power profiles for those cases where the type I error rates appear to be close to the nominal $\alpha$ . Figure 5.1 and Appendix C, Figure C.3, plot estimated powers, denoted $\hat{\kappa}$ , and compare these five tests with small, medium, and large sample sizes for some parameter settings. Using the variance of a binomial distribution, standard errors of these entries $\hat{\kappa}$ are at most 0.021 for 0< $\hat{\kappa} \le$ 0.10 or .90 $\le \hat{\kappa}$ < 1; 0.028 for 0.10< $\hat{\kappa} \le$ 0.20 or .80 $\le \hat{\kappa}$ < 0.90; 0.032 for 0.20< $\hat{\kappa} \le$ 0.30 or .70 $\le \hat{\kappa}$ < 0.80; 0.035 for 0.30< $\hat{\kappa} \le$ 0.70. Furthermore, in Appendix C , Table C.3, I present the results of Cochran's test for testing the equality of the powers among the six tests at fixed alternatives.

$H_0$: The powers are equal.

$H_a$: At least two powers differ.

Cochran's test is used to compare proportions, since all five tests are performed on the same data set.

Summary of power comparisons:

*(1)* When sample sizes are large, or even medium, the power comparison results indicate that all five tests exhibit similar behavior. Further, Cochran's test indicates that the powers only differ for some values of $\pi$, for a few medium and most small samples.

*(2)* The *Plug* p-value test exhibits the highest power when sample sizes are small and medium.

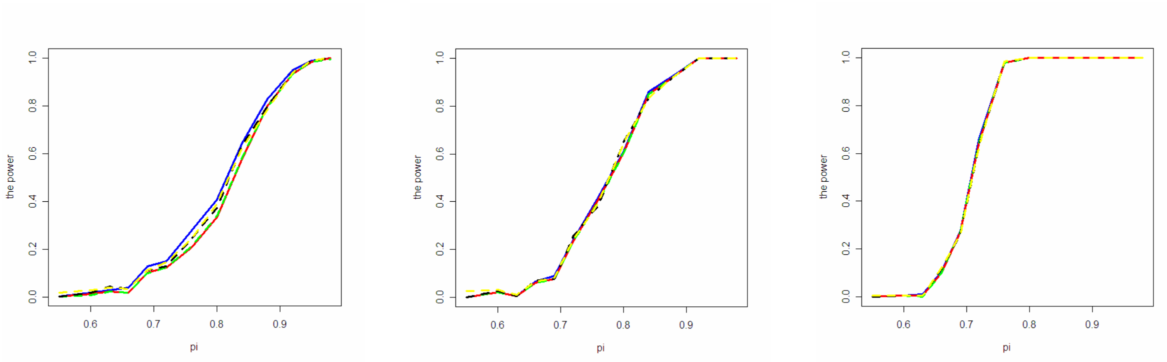
*(3)* The *PBp* and *PBL* p-values have almost the same power. In some small samples, these two p-values-based tests appear to be almost as powerful as the *Plug* p-value test and they appear to be more powerful than the other two, *A_T and A_Z*, in most cases.

*(4)* Clearly, for small or medium samples, the *A_T, and A_Z* have almost exactly the same power. In most small samples, they appear to be less powerful than the others. But they appear to be more powerful than the other two (*PBp* and *PBL*) in some cases.

Overall, from the simulation results for both the level and the power, we conclude that when samples from normal distributions are large or medium, it does not make any practical difference which of these five tests is used. But, because the *PBp* and *PBL* tests require much more time than the others, I recommend using the *plug-in* and the *average p-value* tests. For small samples, I recommend using the *plug-in* test and using *parametric boostrap* tests and *average p-value* tests if controlling the type I error rate is very important. Overall, the *plug-in* test is recommended for all cases in practical applications.

Figure 5.1 Power Simulation of Tests for $IS_{LIN}(\underline{F})$ From Normally Distributed Data

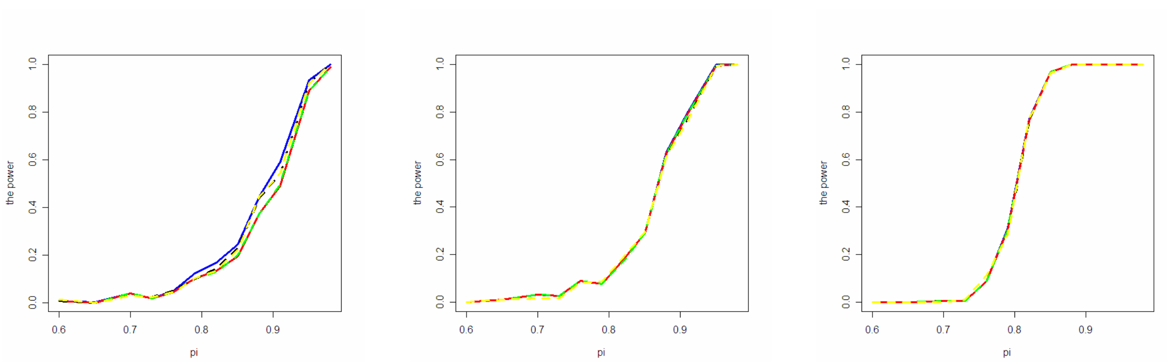
Figure 5.1.1 $K$=3, (a) $(\mu_2) = (0.5)$, $\pi = 0.65$, $\alpha = 0.05$, $\underline{l}^T = (-1/2, -1/2, 1)$




(1) small samples (2) medium samples (3) large samples

Blue: *Plug_in*
Red : *Average_T*
Green : *Average_Z*
Black: *PBT_plugin*
Yellow: *PBT_LRT*

Figure 5.1.2 $K$=3, (b) $(\mu_2, \mu_3) = (0.5, 2)$, $\pi = 0.75$, $\underline{l}^T = (-1/2, -1/2, 1)$




(1) small samples (2) medium samples (3) large samples

Blue: *Plug_in*
Red : *Average_T*
Green : *Average_Z*
Black: *PBT_plugin*
Yellow: *PBT_LRT*

## 5.2 Simulation Study for $IS_{AV}(\underline{F})$ and $IS_{MAX}(\underline{F})$

Now, let us investigate the nonparametric tests for $IS_{AV}(\underline{F})$ and $IS_{MAX}(\underline{F})$. As I already mentioned in Chapter 4, developing exact tests for (4.3) and (4.4) is not possible in general. Instead, I developed tests based on the bootstrap and studied their properties in terms of size and power. This sub-section presents estimated Type I error rates and powers for the nonparametric bootstrap tests (4.3) and (4.4) for $IS_{AV}(\underline{F})$ and $IS_{MAX}(\underline{F})$ .

I first constructed a simple nonparametric bootstrap CI for $IS_{MAX}$ and $IS_{AV}$. Since preliminary simulations indicated that these confidence intervals when used as tests do not work well in most of cases, I switched to CI's based on a pre-pivoting bootstrap (Beran, (1987)) to correct bias and found that it performed better. When a pivot does not exist for a bootstrap test, its actual type 1 error rate could be far from its nominal value, even for large samples. The Prepivoting method was developed by Beran (1987) to produce a bootstrap confidence set that has close to its nominal coverage rate. Beran stated that "Prepivoting is the transformation of a confidence set root by its estimated bootstrap cumulative distribution function". The Prepivoting algorithm for constructing a nonparametric bootstrap CI is given below:

**<u>Algorithm 5.1</u>**: Prepivoting Algorithm for Bootstrap CI's:

Let $T$ be a test statistic, $F$ be the true cumulative distribution function of the data and $\hat{F}_n$ is the empirical cumulative distribution function obtained from the data. Compute the test statistic $T(\hat{F}_n)$ obtained from the data.

(A) Let $\{\underline{y}_1^*, \underline{y}_2^*, ..., \underline{y}_M^*\}$ be independent bootstrap samples drawn from $\hat{F}_n$, Let $\hat{F}_{n,j}^*$ be an estimate of $\hat{F}_n$, obtained from $\underline{y}_j^*$, $j = 1,2,...,M$ . Then compute test statistic $T(\hat{F}_{n,j}^*)$ for each resample $\underline{y}_j^*$, $j = 1,2,...,M$ and the corresponding error term $R_{n,j} = T(\hat{F}_{n,j}^*) - T(\hat{F}_n)$, for $j = 1,2,...,M$ . The empirical cumulative distribution function of the value $\{R_{n,j} : 1 \le j \le M\}$ well approximates $\hat{H}_n$ for sufficiently large $M$, where $\hat{H}_n(x) = \hat{H}_n(x, \hat{F}_n) = pr\{R_n < x \mid \hat{F}_n\}$.

(B) Let $\{\underline{y}^{**}_{j,1}, \underline{y}^{**}_{j,2}, ..., \underline{y}^{**}_{j,B}\}$ be $B$ independent bootstrap samples from $\hat{F}^*_{n,j}$, $j = 1,2,...,M$. Then, compute test statistic $T(\hat{F}^{**}_{j,k})$ for each double resample $\underline{y}^{**}_{j,k}$, $j = 1,2,...,M$, $k = 1,2,...,B$ and the corresponding error terms $R_{n,j,k} = T(\hat{F}^{**}_{j,k}) - T(\hat{F}^*_{n,j})$, for $j = 1,2,...,M$, $k = 1,2,...,B$.

(C) Compute $Z_j = \dfrac{\# R_{n,j,k} \le R_{n,j}}{B}$, for $j = 1,2,...,M$. Then, the empirical cumulative distribution function of the $\{ Z_j : 1 \le j \le M \}$ approximates $\hat{H}_{n,1}$ for sufficiently large $M$ and $B$, where $\hat{H}_{n,1}(x) = H_{n,1}(x, \hat{F}_n) = pr[pr\{R_{n,j,k} < R_{n,j}\} \mid \hat{F}^*_n\} < x \mid \hat{F}_n]$.

(D) An approximate $1-\alpha$, one-sided C.I. for T is given by:

$$
\begin{aligned}
CI_{n,1} &= \{t, R_n(t) \le \hat{H}_n^{-1}\{\hat{H}_{n,1}^{-1}(1-\alpha)\}\} \\
&= \{\hat{T} - T \le \hat{H}_n^{-1}\{\hat{H}_{n,1}^{-1}(1-\alpha)\}\}
\end{aligned}
$$

**Parameter Settings:**

The parameter settings for the simulation study of $IS_{AV}$(F ) and $IS_{MAX}$(F) for Normal distributions are the same as for $IS_{LN}$ (F) with additional choices for pre-pivoting bootstrap resamples. Specifically, set:

$M$=100 ($M$ represents the numbers of the bootstrap resample),
$B$=100 ($B$ is the double bootstrap resample numbers)
Sample sizes: small(10), medium(20), large(50)
$\alpha$ =0.05, 0.10

Time constraints only allowed 100 bootstrap resamples and the simulation results are not good and precise enough to investigate the attained Type I error rates and powers for $IS_{AV}(\underline{F})$ and $IS_{MAX}(\underline{F})$ when $\alpha$ =0.05. Usually, for the more precise results, we need at least 999 bootstrap resamples for $\alpha$ =0.05. So, I only show the results for $\alpha$ =0.10 here.

### 5.2.1 Simulated Type I Error Rates for $IS_{AV}(\underline{F})$

Table 5.3 records estimated Type I error rates of $IS_{AV}(\underline{F})$ obtained by using a prepivoting nonparametric bootstrap CI method. Cells in the table where $\alpha$ = 0.10 does not lie in the approximate .95 confidence interval $\hat{\alpha} \pm 1.96\sqrt{\hat{\alpha}(1-\hat{\alpha})/200}$ are highlighted, where $\hat{\alpha}$ represents the estimated type I error rate. The gray color indicates that the corresponding entry is smaller than the lower bound of the approximate .95 confidence interval. And the pink color indicates that the corresponding entry is greater than the upper bound of this CI.

The error rates in Table 5.3 are close to 0.10 in general even for some small samples, except for the case $\mu_1 = \mu_2 = 0$. That shows us that when two or more populations are very close to each other, the error rates are inflated even for large samples ($n$=50). The, largest inflation might be double what value it should be ($\alpha$ =0.10) when $\pi$ =0.80.

Hence, the prepivoting method seems to work well for $IS_{AV}(\underline{F})$ when two or more distributions are not identical and are fairly far apart.

### 5.2.2 Simulated Type I Error Rates for $IS_{MAX}(\underline{F})$

From the simulation results for the type I error rates of $IS_{MAX}(\underline{F})$ (given in Table 5.4), we notice that when we increase the sample size $n$ to 50, the type I error rate is very close to 0.10, even when $\pi$ is large. So, the prepivoting method for $IS_{MAX}(\underline{F})$ appears to work well for large sample sizes.

But for the small samples, the Type I error rates exhibit serious inflation and even can be as large as 0.49 when $\pi = 0.95$. So for $IS_{MAX}(\underline{F})$, the pre-pivoting method is not useful for small samples.

Therefore, for $IS_{MAX}(\underline{F})$, the simulation studies show that the pre-pivoting method is helpful in reducing the bias for some cases, especially for the large samples.

For both $IS_{AV}(\underline{F})$ and $IS_{MAX}(\underline{F})$, there is some inflation of estimated Type I error rates. I will investigate this issue in future research.

Table 5.3 Prepivoting Nonparametric Bootstrap CI for $IS_{AV}(\underline{F})$

Table 5.3.1 $K$=3, $\alpha$=0.10

| Type I error ($\alpha$) | **(a) ($\mu_2$=0)** | | | **(a) ($\mu_2$=0.5)** | | | **(a) ($\mu_2$=1)** | | |
|---|---|---|---|---|---|---|---|---|---|
| | $\pi$ | | | $\pi$ | | | $\pi$ | | |
| | 0.60 | 0.70 | 0.80 | 0.60 | 0.70 | 0.80 | 0.70 | 0.80 | 0.85 |
| $n$=(10,8,9) | **0.19** | **0.21** | **0.16** | **0.15** | **0.13** | **0.13** | **0.13** | **0.08** | **0.1** |
| $n$=(20,25,28) | **0.2** | **0.19** | **0.22** | **0.11** | **0.14** | **0.12** | **0.13** | **0.11** | **0.1** |
| $n$=(50,60,55) | **0.18** | **0.14** | **0.25** | **0.11** | **0.09** | **0.06** | **0.1** | **0.08** | **0.1** |

| Type I error ($\alpha$) | **(b) $(\mu_2, \mu_3)$ = (0.5, 1)** | | **(b) $(\mu_2, \mu_3)$ = (0.5, 2)** | | | **(b) $(\mu_2, \mu_3)$ = (0.5, 5)** | | |
|---|---|---|---|---|---|---|---|---|
| | $\pi$ | | $\pi$ | | | $\pi$ | | |
| | 0.65 | 0.70 | 0.65 | 0.75 | 0.80 | 0.65 | 0.75 | 0.85 |
| $n$=(10,8,9) | **0.08** | **0.13** | **0.07** | **0.14** | **0.12** | **0.05** | **0.14** | **0.09** |
| $n$=(20,25,28) | **0.12** | **0.12** | **0.10** | **0.14** | **0.08** | **0.1** | **0.1** | **0.05** |
| $n$=(50,60,55) | **0.08** | **0.14** | **0.16** | **0.09** | **0.13** | **0.14** | **0.12** | **0.10** |

Table 5.3.2 $K$=5, $\alpha$=0.10

| Type I error ($\alpha$) | **(a) (0,0,1)** | | | **(a) (0.5, 0.5, 1)** | | | **(b) (0, 0, 2, 2)** | | **(b) (0.5, 1, 3, 3.5)** | |
|---|---|---|---|---|---|---|---|---|---|---|
| | $\pi$ | | | $\pi$ | | | $\pi$ | | $\pi$ | |
| | 0.66 | 0.70 | 0.75 | 0.65 | 0.70 | 0.75 | 0.71 | 0.74 | 0.77 | 0.80 |
| $n$=(10,8,9,7,9) | **0.24** | **0.13** | **0.17** | **0.16** | **0.15** | **0.15** | **0.22** | **0.30** | **0.09** | **0.12** |
| $n$=(20,25,22,24,26) | **0.28** | **0.17** | **0.28** | **0.09** | **0.13** | **0.15** | **0.17** | **0.26** | **0.06** | **0.14** |
| $n$=(50,60,55,54,58) | **0.17** | **0.12** | **0.16** | **0.16** | **0.19** | **0.15** | **0.23** | **0.24** | **0.13** | **0.14** |

Table 5.4 Prepivoting Nonparametric Bootstrap CI for $IS_{MAX}(\underline{F})$

Table 5.4.1 $K$=3, $\alpha$ =0.10

| Type I error ($\alpha$) | (a) $\mu_2$=0 | | | | (a) $\mu_2$=0.5 | | | |
|---|---|---|---|---|---|---|---|---|
| | $\pi$ | | | | $\pi$ | | | |
| | 0.65 | 0.75 | 0.85 | 0.95 | 0.65 | 0.75 | 0.85 | 0.95 |
| $n$=(10,8,9) | **0.13** | **0.23** | **0.32** | **0.49** | **0.12** | **0.14** | **0.26** | **0.47** |
| $n$=(20,25,28) | **0.16** | **0.12** | **0.14** | **0.23** | **0.11** | **0.09** | **0.09** | **0.23** |
| $n$=(50,60,55) | **0.16** | **0.11** | **0.15** | **0.13** | **0.09** | **0.11** | **0.10** | **0.11** |

Table 5.4.2 $K$=3, $\alpha$ =0.10

| Type I error ($\alpha$) | (b) $(\mu_2, \mu_3)$ = (0.5, 1) | | | | (b) $(\mu_2, \mu_3)$ = (0.5, 2) | | | |
|---|---|---|---|---|---|---|---|---|
| | $\pi$ | | | | $\pi$ | | | |
| | 0.65 | 0.70 | 0.75 | 0.80 | 0.65 | 0.75 | 0.85 | 0.95 |
| $n$=(10,8,9) | **0.16** | **0.12** | **0.14** | **0.13** | **0.11** | **0.1** | **0.26** | **0.42** |
| $n$=(20,25,28) | **0.09** | **0.05** | **0.05** | **0.11** | **0.08** | **0.08** | **0.12** | **0.19** |
| $n$=(50,60,55) | **0.18** | **0.09** | **0.14** | **0.09** | **0.16** | **0.08** | **0.15** | **0.09** |

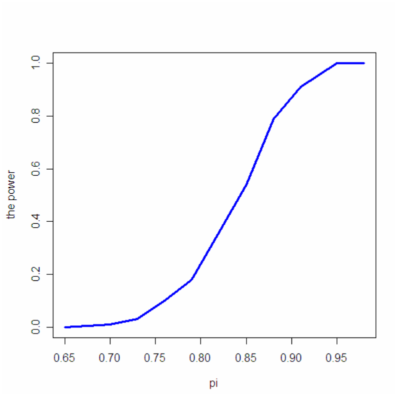
Table 5.4.3 K=5, $\alpha$ =0.10

| Type I error ($\alpha$) | (a) $(\mu_2, \mu_3, \mu_4)$ = (0,0,1) | | | | (a) $(\mu_2, \mu_3, \mu_4)$ = (0.5,0.5,1) | | | |
|---|---|---|---|---|---|---|---|---|
| | $\pi$ | | | | $\pi$ | | | |
| | 0.80 | 0.85 | 0.90 | 0.95 | 0.80 | 0.85 | 0.90 | 0.95 |
| $n$=(10,8,9,7,9) | **0.43** | **0.53** | **0.61** | **0.69** | **0.32** | **0.32** | **0.36** | **0.54** |
| $n$=(20,25,22,24,26) | **0.21** | **0.21** | **0.33** | **0.47** | **0.11** | **0.07** | **0.1** | **0.18** |
| $n$=(50,60,55,54,58) | **0.21** | **0.22** | **0.15** | **0.18** | **0.08** | **0.10** | **0.09** | **0.05** |

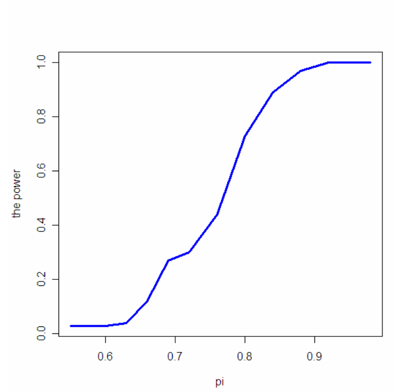
### 5.2.3 Simulated Powers for $IS_{AV}(\underline{F})$ and $IS_{MAX}(\underline{F})$

I also investigated the power properties of tests for both $IS_{AV}(\underline{F})$ and $IS_{MAX}(\underline{F})$. When the attained Type I error rate is close to its nominal value ($\alpha$ =0.10), the simulated-powers results for $IS_{AV}(\underline{F})$ (illustrated in Appendix C, Figure C.4) and for $IS_{MAX}(\underline{F})$ (given in Figure 5.2) indicate that the prepivoting Bootstrap tests have an increasing power function when the value of $\pi$ increases.

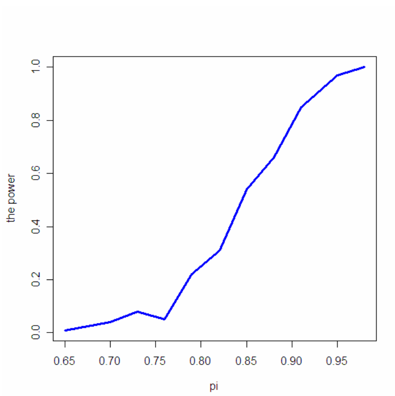
Figure 5.2 Power Simulation Results for $IS_{MAX}(\underline{F})$, Medium Samples



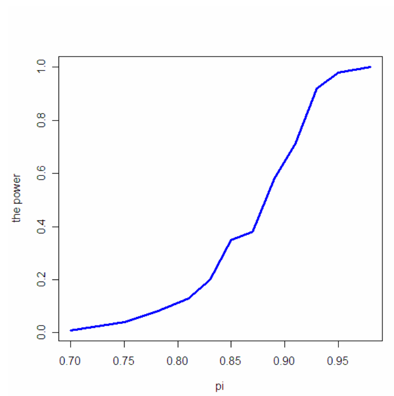

(a) $K$=3 (a) $(\mu_2) = (0.5)$, $\pi = 0.75$




(b) $K$=3 (b) $(\mu_2, \mu_3) = (0.5, 1)$, $\pi = 0.65$




(c) $K$=3 (b) $(\mu_2, \mu_3) = (0.5, 2)$,
$\pi = 0.75$




(d) $K$=5, (a) $(\mu_2, \mu_3, \mu_4) = (0.5, 0.5, 1)$,
$\pi = 0.80$,

## 5.3 Example

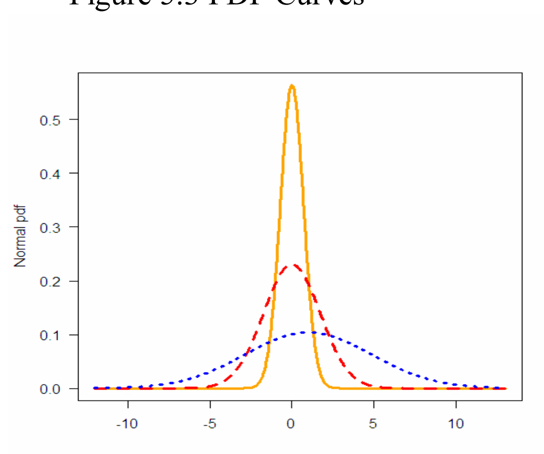
Consider three Normal distributions:

$X_1 \sim N(0, \sigma_1^2 = 0.5)$,

$X_2 \sim N(0, \sigma_2^2 = 3)$,

$X_3 \sim N(1, \sigma_3^2 = 14.705)$.

Figure 5.3 PDF Curves



Although the means are not identical, there is considerable overlap among the three distributions, as pictured in Figure 5.3 and as indicated by the following values of some intrinsic separation parameters:

$$IS_{LIN}(\underline{F}) = 0.2533 \Rightarrow \ \Phi(IS_{LIN}(\underline{F})) = 0.60. \ \text{for } \underline{l}^{T} = \{l_1 = l_2 = -1/2, \ l_3 = 1\}$$

$$IS_{AV}(\underline{F}) = 0.5650, \qquad IS_{MAX}(\underline{F}) = 0.6012.$$

We generated independent random samples of three sample sizes, small, medium, and large, from these three distributions. Summary statistics are listed in Table 5.5 and the corresponding side by side boxplots are given in Figure 5.4. The considerable overlap among these boxplots is, of course, what decision-makers using these data sets would see. The five p-values for testing the value of $IS_{LIN}$ using the hypothesis in (2.4) with $\pi = 0.50$, 0.55, 0.60, 0.65, 0.70 are given in Table 5.6.

First, note that all tests yield results which support the conclusion that the three means are not identical, i.e., $\pi > 0.50$. As $\pi$ increases, indicating increasing separation, the tests provide increasing support for the hypothesis that the distributions are not 'far' apart, a main point of this dissertation. Except for 'small' samples, the p-values of all the tests are very similar. This example motivates me to investigate in the future procedures for selecting sample sizes so that my tests have desired power at specified alternatives.

Table 5.5 Summary Statistics of the Data Sets for the Example

| Treatments | $\mu_i$ | $\sigma_i^2$ | $n_i$ | $\bar{x}_i$ | $s_i^2$ |
|---|---|---|---|---|---|
| $X_1$ | 0 | 0.5 | 10 | -0.0251 | 0.3745 |
| | | | 30 | 0.0244 | 0.4039 |
| | | | 100 | -0.0433 | 0.5370 |
| $X_2$ | 0 | 3 | 15 | 0.0714 | 1.705 |
| | | | 50 | 0.1058 | 3.3723 |
| | | | 150 | -0.1622 | 2.9977 |
| $X_3$ | 1 | 14.7050 | 12 | 2.7141 | 10.2982 |
| | | | 20 | 2.5142 | 16.0952 |
| | | | 120 | 1.3517 | 15.1650 |

Figure 5.4 Side by Side Boxplots for the Data Sets

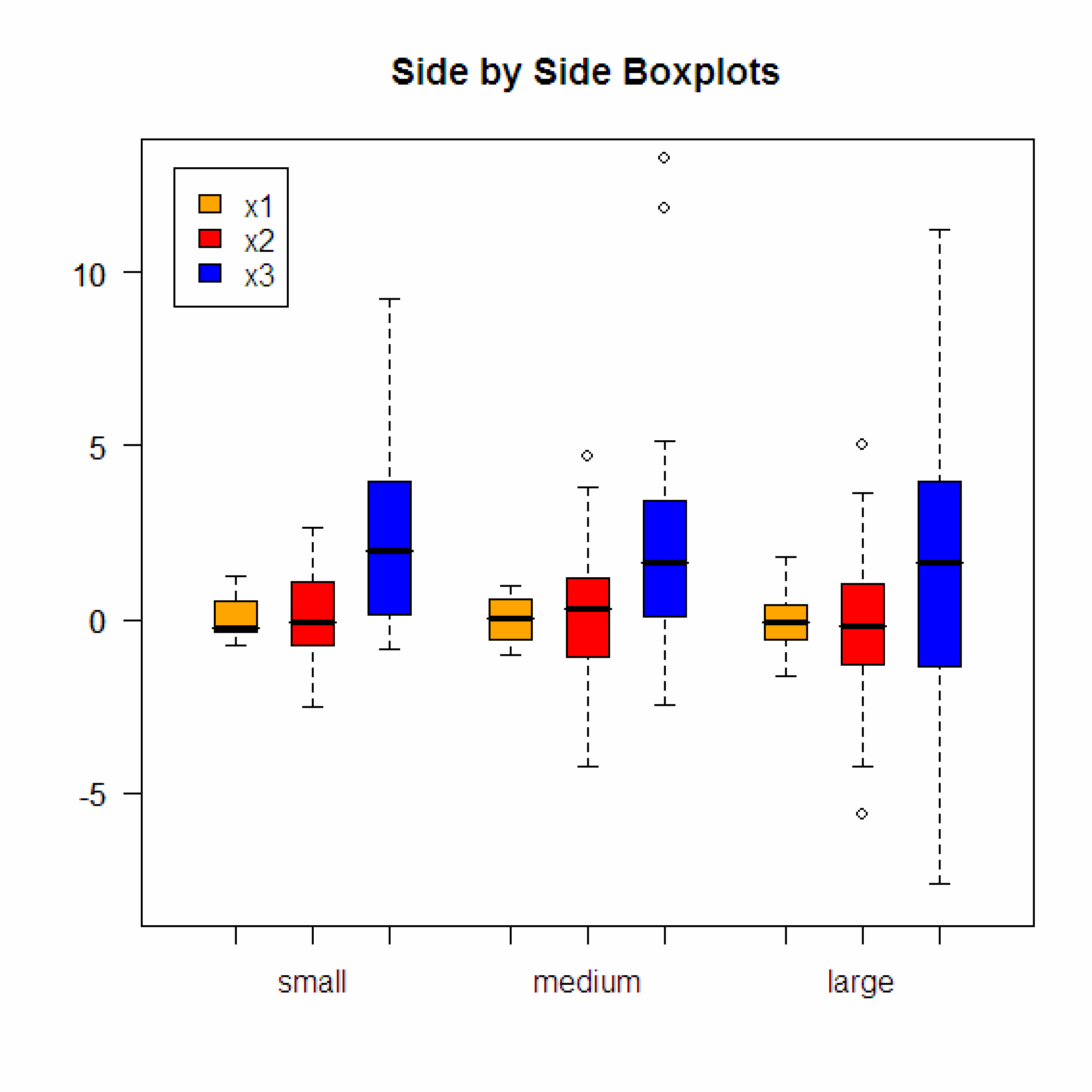



Table 5.6 Comparison of the P-Values for $IS_{LIN}(\underline{F})$ for the Example

Table 5.6.1 $\pi = 0.50$

| Sample Sizes | ***Plug*** | ***A_T*** | ***A_Z*** | ***PBp*** | ***PBL*** |
|---|---|---|---|---|---|
| Small | 0.0074 | 0.0079 | 0.0081 | 0.010 | 0.0227 |
| Medium | 0.0069 | 0.0075 | 0.0075 | 0.010 | 0.0222 |
| Large | 5.4168e-05 | 5.4151e-05 | 5.2923e-05 | 0.010 | 0.0128 |

Table 5.6.2 $\pi$ = 0.55

| Sample Sizes | ***Plug*** | ***A_T*** | ***A_Z*** | ***PBp*** | ***PBL*** |
|---|---|---|---|---|---|
| Small | 0.0202 | 0.020 | 0.0193 | 0.050 | 0.050 |
| Medium | 0.0268 | 0.0272 | 0.0269 | 0.020 | 0.020 |
| Large | 0.0057 | 0.0058 | 0.0059 | 0.020 | 0.020 |

Table 5.6.3 $\pi$ = 0.60

| Sample Sizes | ***Plug*** | ***A_T*** | ***A_Z*** | ***PBp*** | ***PBL*** |
|---|---|---|---|---|---|
| Small | 0.0486 | 0.0496 | 0.0508 | 0.090 | 0.100 |
| Medium | 0.0818 | 0.0835 | 0.0825 | 0.140 | 0.140 |
| Large | 0.1222 | 0.1216 | 0.1218 | 0.140 | 0.140 |

Table 5.6.4 $\pi$ = 0.65

| Sample Sizes | ***Plug*** | ***A_T*** | ***A_Z*** | ***PBp*** | ***PBL*** |
|---|---|---|---|---|---|
| Small | 0.1046 | 0.1075 | 0.1081 | 0.070 | 0.080 |
| Medium | 0.2007 | 0.2013 | 0.1981 | 0.220 | 0.220 |
| Large | 0.5969 | 0.6013 | 0.6025 | 0.590 | 0.750 |

Table 5.6.5 $\pi$ = 0.70

| Sample Sizes | ***Plug*** | ***A_T*** | ***A_Z*** | ***PBp*** | ***PBL*** |
|---|---|---|---|---|---|
| Small | 0.2025 | 0.2050 | 0.2056 | 0.220 | 0.230 |
| Medium | 0.3990 | 0.4045 | 0.4048 | 0.440 | 0.440 |
| Large | 0.9582 | 0.9589 | 0.9594 | 0.90 | 0.77 |

Table 5.7 exhibits the lower limits ($\hat{L}(Data)$) of one-sided lower confidence sets for $IS_{AV}(\underline{F})$ and $IS_{MAX}(\underline{F})$ for the data in this example, obtained by using a nonparametric bootstrap and prepivoting. We would conclude that the separation among the distributions, as measured by $IS_{AV}$ or $IS_{MAX}$, is greater than $\pi$ only if $\pi < \hat{L}(Data)$. Again, as $\hat{L}(Data)$ increases, indicating increasing separation, the tests provide decreasing support for the hypothesis that the distributions are 'that far' apart. The dividing line here between separation and not separation is for $\pi$ around 0.56 for $IS_{AV}$ and around 0.60 for $IS_{MAX}$. We note that these bootstrap results depend, hopefully weakly, on the resample numbers ($M$, $B$).

Table 5.7 Lower Limit of CI for $IS_{AV}(\underline{F})$ and $IS_{MAX}(\underline{F})$ for the Example

Table 5.7.1 Lower Limit of CI for $IS_{AV}(\underline{F})$

| Sample Sizes | $\alpha = 0.05$ | | $\alpha = 0.10$ | |
|---|---|---|---|---|
| | *M=B=100* | *M=B=500* | *M=B=100* | *M=B=500* |
| Small | 0.5489 | 0.5466 | 0.5553 | 0.5719 |
| Medium | 0.5743 | 0.5348 | 0.5799 | 0.5558 |
| Large | 0.5525 | 0.5559 | 0.5544 | 0.5670 |

Table 5.7.2 Lower Limit of CI for $IS_{MAX}(\underline{F})$

| Sample Sizes | $\alpha = 0.05$ | | $\alpha = 0.10$ | |
|---|---|---|---|---|
| | *M=B=100* | *M=B=500* | *M=B=100* | *M=B=500* |
| Small | 0.6167 | 0.60 | 0.6167 | 0.60 |
| Medium | 0.525 | 0.5683 | 0.6267 | 0.6364 |
| Large | 0.5792 | 0.5692 | 0.5809 | 0.5804 |

# CHAPTER 6 Summary and Conclusion

I developed and explored the concept and tests for some Intrinsic Separation Parameters: $IS_{LIN}(\underline{F})$, $IS_{MAX}(\underline{F})$, and $IS_{AV}(\underline{F})$), assuming normality for $IS_{LIN}(\underline{F})$ , among two or more distributions which may have different shapes by using frequentist, Bayesian, Fiducial and bootstrap modes of inference . Over all, the tests developed for ISP among normal distributions with unequal variances are more complex than the one-way ANOVA, which tests the equality of means with equal variances.

For $IS_{LIN}(\underline{F})$, I developed five tests for this ISP, assuming the normality. They are the *plug-in test (Plug), two average p-value tests (A_T, A_Z), and two parametric boostrap tests (PBp, PBL)*. The asymptotic behavior of the parametric bootstrap test and the average p-value tests are derived in Chapter 2 and Chapter 3 separately. Chapter 2 presents a method for proving that the parametric bootstrap test for the LRT is an asymptotically size-$\alpha$ test under normality and some mild conditions. Chapter 3 proves that the average p-value ( $\bar{p}$ ) is *consistent* under normality. In addition, based on the simulation results in Chapter 5, in terms of estimated size and power, my five testing procedures perform very similarly and very well for medium and large samples. Furthermore, Example 5.3 illustrates these five tests and shows them to behave similarly. In general, it does not matter which of these five procedures for medium and large samples ($\geq$30), we use.

When samples are very small ( $\leq 15$), simulation results show that the *plug-in test* performs well regardless of the values of the error variances, and the number of distributions being compared, except it has type I error rates inflated for a few cases. Other tests, compared to the *plug-in test*, are more conservative, but with the loss of power (significant in most of cases).

Since a meaningful interpretation of $IS_{LIN}(\underline{F})$ depends heavily on the assumption of normality, the issue of robustness of tests for it is of limited interest. Defining and investigating more ISP's such as $IS_{MAX}(\underline{F})$ and $IS_{AV}(\underline{F})$) which do not depend on the form of the underlying distributions would be an important step forward. Constructing effective tests for ISP for $IS_{AV}(\underline{F})$ and $IS_{MAX}(\underline{F})$ remains a challenging problem. Prepivoting a bootstrap to reduce the bias in constructing a one-sided lower confidence set for $IS_{AV}(\underline{F})$ and $IS_{MAX}(\underline{F})$ only appears to works in some cases. This issue warrants further study.

# CHAPTER 7 Further Researches

(1) Investigate robustness of my tests with respect to the presence of outliers.

(2) Develop procedures for selecting sample sizes so that my tests have desired power at specified alternatives.

(3) Develop another approach to compare several distributions (sometimes called counting overlap), which is an intrinsic separation test based on the proportion of overlapping observations.

(4) Further explore the concept of separation for skewed families of location scale distributions such as the extreme value.

(5) Investigate the ISP for $IS_{AV}(\underline{F})$ and $IS_{MAX}(\underline{F})$ by using BCa (Bias Corrected and Accelerated) method or develop other procedures.

(6) Derive the asymptotic distribution of the likelihood ratio test for $IS_{LIN}(\underline{F})$.

## References


- Aspin, A. A. (1948) "An examination and further development of a formula arising in the problem of comparing two mean values" *Biometrika*, 35, 88-96. (Extended Welch's series expansion solution to the fourth degree in the reciprocal of the degrees of freedom.)
- Bayarri, M.J. and Berger, J.O.(2000) "P Values for Composite Null Models", *Journal of the American Statistical Association*, 95, 1127-1142.
- Barnard, G. A. (1984), "Comparing the Means of Two Independent Samples," *Appl.Statist.* , 33, No. 3, 266-271.
- Beran, R. (1987), "Prepivoting to reduce level error of confidence sets", *Biometrika*, 74, 3, 457-468.
- Beran, J. and Ducharme, G.(1991) "Asymptotic Theory for Bootstrap Methods in Statistics",*Center de Recherches Mathematiques*, University of Montreal.
- Berger, J.O. (1985), *Statistical Decision Theory and Bayesian Analysis (2nd ed.)*, New York: Springer-Verlag.
- Berger, J.O. and Selke, T.(1987), "Testing a Point Null Hypothesis: The Irreconcilability of P Values and Evidence", *Journal of the American Statistical Association*, 82, 112-122.
- Berger, R.L. and Boos, D. (1994) "P Values Maximized Over a Confidence Set for the Nuisance Parameter" *Journal of the American Statistical Association*, 89, 1012-1016.
- Buot, M.L.G, Hosten, S. and Richards, D. St. P. (2007) "Counting and Locating the Solutions of Polynomial Systems of Maximum Likelihood Equations, II: the Behrens-Fisher Problem", *Statistica Sinica*, **17**, 1343-1354.
- Cochran, W.G. (1941). "The distribution of the largest of a set of estimated variances as a fraction of their total." *Ann. Eugenics*, 11 47-52
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences (2nd ed)*, Hillsdale, NJ: Lawrence Erlbaum Associates.

- Delaney, H. D., and Vargha, A.(2002), "Comparing Several Robust Tests of Stochastic Equality with Ordinally Scaled Variables and Small to Moderate Sized Samples" *Psychological Methods*, 7, 485-503.
- Drton, M. (2008) "Multiple Solutions to the Likelihood Equations in the Behrens-Fisher Problem", *Statist.&Prob. Letters*, 78(18), 3288-3293.
- Drton, M. (2009) ' Likelihood Ratio Tests and Singularities,' *Annals of Statistics*, 37, 979-1012.
- Doornbos, R., and Prins, H.J., (1956) "Slippage tests for a set of gamma variates" *Indgationes Mathematicae*, 18, 329-337.
- Doornbos, R. (1966), *Slippage Tests*, Math. Centre Tracts No. 15, Mathematisch Centrum, Amsterdam.
- Edwards, , A.W.F. (1997) " What Did Fisher Mean by Inverse Probability in 1912-1922?", *Statistical Science*, 12, 177-184.
- Efron, B. (1987), "Better Bootstrap Confidence Intervals" (with discussion), *Journal of the American Statistical Association*, 82, 171-200.
- Efron, B. & Tibshirani, R.J. (1993), *An Introduction to the Bootstrap*, Chapman & Hall.
- Fisher, R. A. (1956), *Statistical Methods and Scientific Inference*, New York: Hafner
- Hall, P. and Martin, M. (1988) "On the Bootstrap and Two-Sample Problems", *Australian J. Statist*., 179-192.
- Hess, M. R., Hogarty, K.R., Ferron, J., and Kromrey, J. (2007) "Interval estimates of multivariate effect sizes", *Educational and Psychological Measurement*, 21-40.
- Hodges, J.L., Jr., and Lehmann, E.L. (1954), "Testing the Approximate Validity of Statistical Hypotheses", *Journal of the Royal Statistical Society. Series B (Methodological)*, 16, 261-268.
- Jaber, K. and Cox, T., "The Generalized likelihood Ratio Test for the Beherns-Fisher Problem", Personal Communication.
- Jonckheere, A. R. (1954) "A Distribution Free k-sample Test against Ordered Alternatives," *Biometrika* 41:133-145.

- Kemp, K.E., Yang, S.S, Perng, S.K., and Nelson, P.I. (1993) "An asymptotically distribution-free test for assessing the separation between two distributions", *Journal of Nonparametric Statistics*, 2, 235-248.
- Krishnamoorthy, K., Lu, F., and Mathew, T. (2007), "A parametric bootstrap approach for ANOVA with unequal variances: Fixed and random models," *Computational Statistics & Data Analysis*, 51, 5731-5742.
- Kudo, A. (1956) "On the invariant multiple decision procedures," *Bull. Math. Stat.*, 6, 57-69
- Lehmann, E.L. and Romano, Joseph P. (2005), *Testing Statistical Hypotheses (Third Edition)*, Springer, NY.
- Mann, H.B., and Whitney, D.R.(1947), "On a Test of Whether One of Two Random Variables is Stochastically Larger Than the Other," *Annals of Mathematical Statistics*, 18,50-60
- Martin, M. A. (2007), "Bootstrap hypothesis testing for some common statistical problems: A critical evaluation of size and power properties." *Computational statistics & Data analysis*, 51 6321-6342
- Meng, X.,(1994), "Posterior Predictive p-Values," *The Annals of Statistics*, 22, 3, 1142-1160
- Mielke, P.W. and Berry, K.J. (1994), "Permutation tests for common locations among samples with unequal variances," *Journal of Educational and Behavioral Statistics*; 19: 217-236
- Mosteller, F.,(1948), "A k-sample slippage test for an extreme population," *Ann, Math. Stat.*, 19, 58-65
- Murphy, K. (1990), " If the null hypothesis is impossible, why test it?", *American Psychologist*, 45, 403-404
- Murphy, K., and Myors, B. (1999), "Testing the hypothesis that treatments have negligible effects: Minimum-effect tests in the general linear model." *Journal of Applied Psychology*, 84, 234-248
- Murphy, K.R. and Myors, B. (2004), *Statistical Power Analysis, Second Edition*, LEA; Mahwah, NJ.

- Nair, Vijayan N. (1982), “Q-Q Plots with Confidence Bands for Comparing Several Populations,” *Scand. J. Statist*, 9:193-200
- Paulson, E., (1952), “An optimum solution to the k-sample slippage problem for the normal distribution,” *Ann. Math. Stat.*, 23, 610-616
- Perng, S.K., Kemp, K.E. and Nelson, P.I. (1989), “Testing for a separation between tw normal distributions”, *Commun. Statist.-Theory Meth.*, 18(5), 1895-1912
- Reiczigel, Jeno, Zakaria’s, Iidiko, and Ro’zsa, Lajos (2005), “A Bootstrap test of Stochastic Equality of Two Populations,” *The American Statistician*, 59, 2, 156-161
- Rice, W.R. and Gaines, S.D. (1989), “One-way analysis of variance with unequal variances,” *Proc. Natl. Acad. Sci. of USA*, 86. 8183-8184
- Robins, J.M., Vaart, A., and Ventura V. (2000), “Asymptotic Distribution of P Values in Composite Null Models”, *Journal of the American Statistical Association*, 95, 1143-1172
- Romano, J. P. (1990), “On the behavior of randomization tests without a group invariance assumption.” *Journal of the American Statistical Association*, 85, 411, 686-692
- Rouanet, H. (1996), “Bayesian methods for assessing importance of effects,” *Psychological Bulletin*, 119, 149-158
- Satterthwaite, F. E. (1946), "An Approximate Distribution of Estimates of Variance Components,", *Biometrics Bulletin 2*: 110-114
- Serlin, R.A., & Lapsley, D.K. (1993), “Rational appraisal of psychological research and the good-enough principle.” In G. Keren & C. Lewis (Eds.), *A handbook for data analysis in the behavioral sciences:Methodological issues* (pp. 199-228). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Silvey, S.D.(1975), *Statistical Inference*, London: Chapman and Hall.
- Terpstra, T. J. (1952), “The Asymptotic Normality and Consistency of Kendall’s Test Against Trend, When Ties are Present in One Ranking.” *Indagationes Math* 14:327-333.
- Tsui, K., and Weerahandi, S. (1989), “Generalized p-Values in Significance Testing of Hypotheses in the Presence of Nuisance Parameters.” *Journal of the American Statistical Association*, 84, 602-607.

- Van der Vaart, A.W. (1998), *Asymptotic Statistics*, Cambridge University Press, Cambridge.
- Welch, B.L.(1938), "The Significance of the Difference Between Two Means When the Population Variances are Unequal," *Biometrika*, 29,350-362
- Welch, B. L. (1947), "The generalization of 'Student's' problems when several different population variances are involved" *Biometrika*, 34, 28-35.
- Wilcoxon, F.(1945), " Individual Comparisons by Ranking Methods," *Biometrics*, 1, 80-83
- Wilson, E.B. (1927), "Probable Inference, the Law of Succession, and Statistical Inference" *Journal of the American Statistical Association*, 22, 209-212.
- Zimmerman, D.W., and Zumbo, B., D.(1993), "Rank Transformations and the Power of the Student t-Test and Welch t-Test for Nonnormal Populations with Unequal Variances," *Canadian Journal of Experimental Psychology*, 47, 523-539.

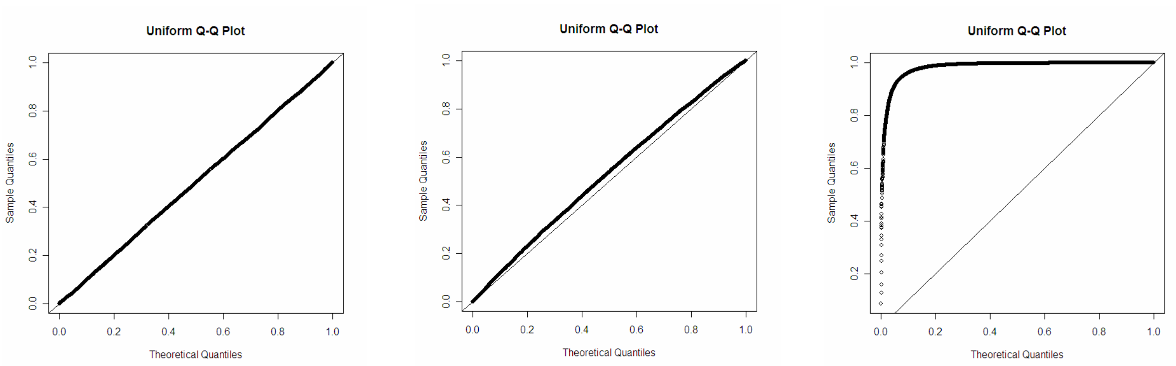
# APPENDIX A

**Figure A.1 Estimated P-Values Under (2.4):** $H_0$ **:** $\mathrm{IS}_{\mathrm{LIN}}(\underline{\mathrm{F}}) \le \eta_\pi$

Satterthwaite Approximation (Conservative) Test—p-value (2.9) , 10000 Iterations

(A.1.1) $X_1 \underset{D}{=} X_2 \sim \mathrm{N}(0,1)$, $X_3 \sim \mathrm{N}(1.56767,1)$, $\underline{l}^T = (-1/2,-1/2,1)$

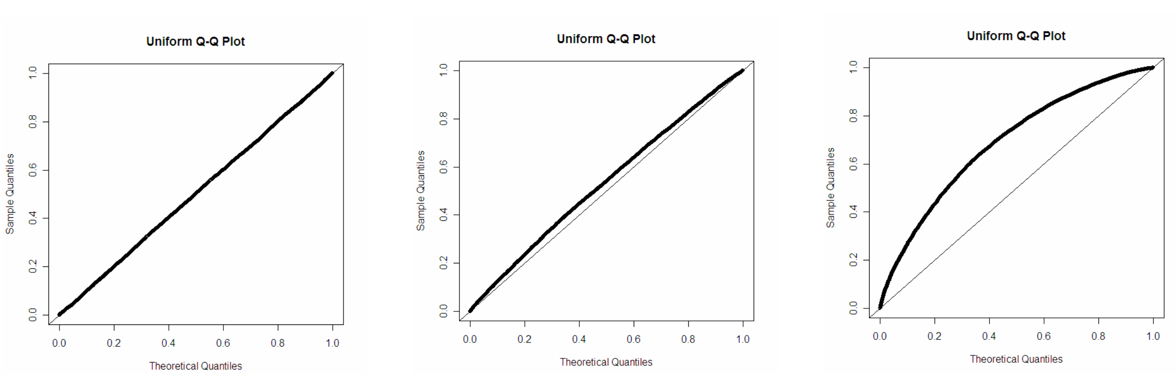
Figure A.1.1 Uniform Q-Q Plot of Estimated P- Value




(1) $n_1 = n_2 = n_3 = 10$ (2) $n_1 = 10,\ n_2 = 11,\ n_3 = 12$ (3) $n_1 = 10,\ n_2 = 30,\ n_3 = 90$

(A.1.2) $X_1 \sim \mathrm{N}(0,1)$, $X_2 \sim \mathrm{N}(0,(\sqrt{10})^2)$, $X_3 \sim \mathrm{N}(3,(\sqrt{2.74316})^2)$, $\underline{l}^T = (-1/2,-1/2,1)$

Figure A.1.2 Uniform Q-Q Plots of Estimated P- Values



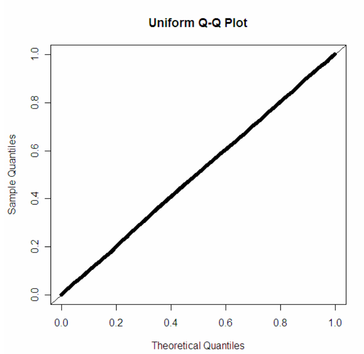

(1) $n_1 = n_2 = n_3 = 10$ (2) $n_1 = 10,\ n_2 = 11,\ n_3 = 12$ (3) $n_1 = 100,\ n_2 = 125,\ n_3 = 150$

**Figure A.2 Estimated P-Values Under $H_0$ (2.4)**

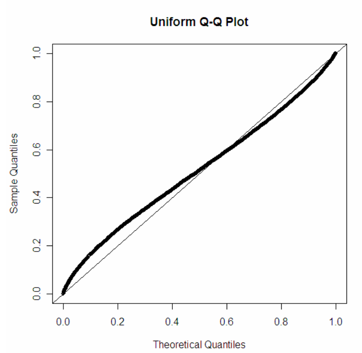
Satterthwaite Approximation (Estimate) test—p-value (2.12)

(A.2.1) $X_1 \underset{D}{=} X_2 \sim \mathrm{N}(0,1)$, $X_3 \sim \mathrm{N}(1.56767,1)$, $\underline{l}^T = (-1/2, -1/2, 1)$, 10000 Iterations

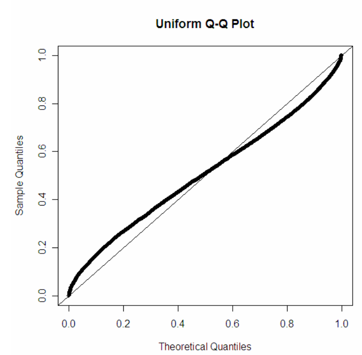
Figure A.2.1 Uniform Q-Q Plots of Estimated P- Values




(1) $n_1$=10, $n_2$=15, $n_3$=20



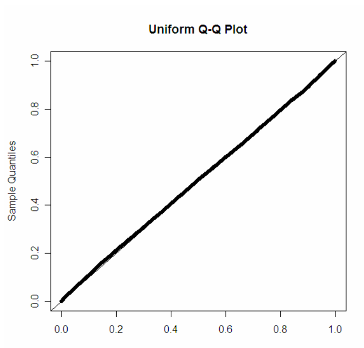

(2) $n_1$=10, $n_2$=30, $n_3$=90




(3) $n_1$=100, $n_2$=500, $n_3$=900

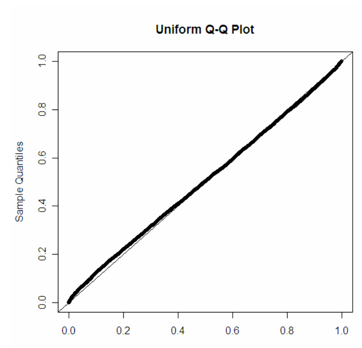
(A.2.2) $X_1 \sim \mathrm{N}(0,1)$, $X_2 \sim \mathrm{N}(0,(\sqrt{10})^2)$, $X_3 \sim \mathrm{N}(3,(\sqrt{2.74316})^2)$, $\underline{l}^T = (-1/2, -1/2, 1)$,
B=10000 Iterations

Figure A.2.2 Uniform Q-Q Plots of Estimated P- Values




(1) $n_1$=10, $n_2$=30, $n_3$=15


(2) $n_1$=10, $n_2$=100, $n_3$=30



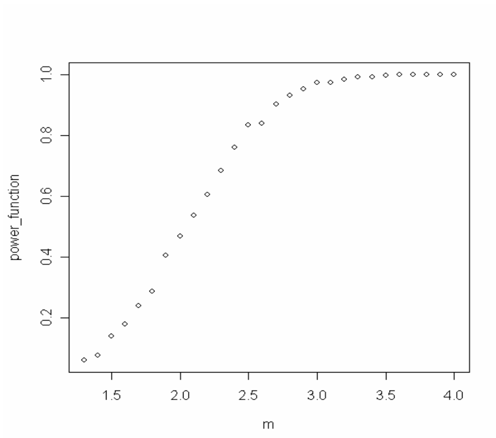

(3) $n_1$=100, $n_2$=1000, $n_3$=300

**Figure A.3 Estimated Power ( $H_0$ :(2.4) )**

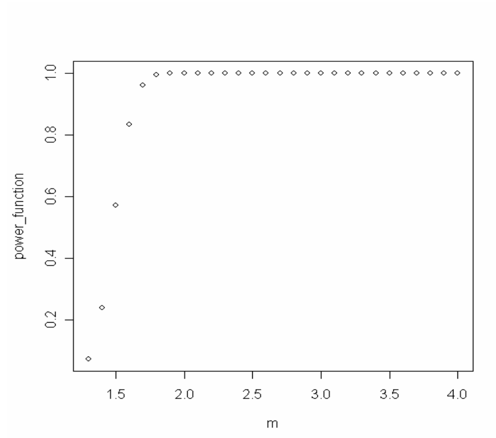
Satterthwaite Approximation (Estimate) Test (p-value (2.12))—Power

(A.3.1) $X_1 \underset{D}{=} X_2 \sim \mathrm{N}(0,1)$, $X_3 \sim \mathrm{N}(1.56767,1)$, $\underset{\sim}{l}^T = (-1/2, -1/2, 1)$, 1000 Iterations, m = $n_\pi$.

Figure A.3.1 Power vs $n_\pi$



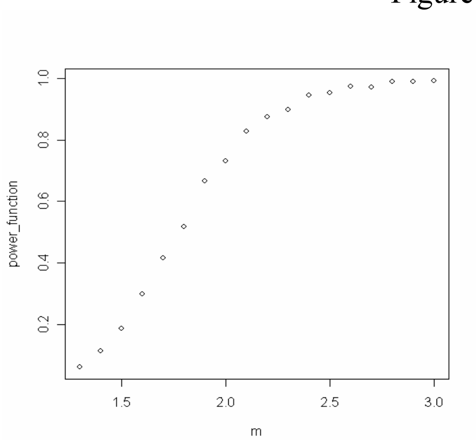

(1) $n_1 = n_2 = n_3 = 10$



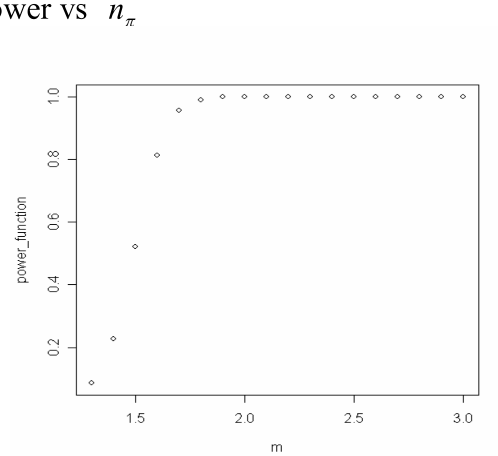

(2) $n_1 = n_2 = n_3 = 100$

(A.3.2) $X_1 \sim \mathrm{N}(0,1)$, $X_2 \sim \mathrm{N}(0, (\sqrt{10})^2)$, $X_3 \sim \mathrm{N}(3, (\sqrt{2.74316})^2)$, $\underset{\sim}{l}^T = (-1/2, -1/2, 1)$, B=1000 Iterations, m = $n_\pi$.

Figure A.3.2 Power vs $n_\pi$




(1) $n_1 = 10, n_2 = 15, n_3 = 25$



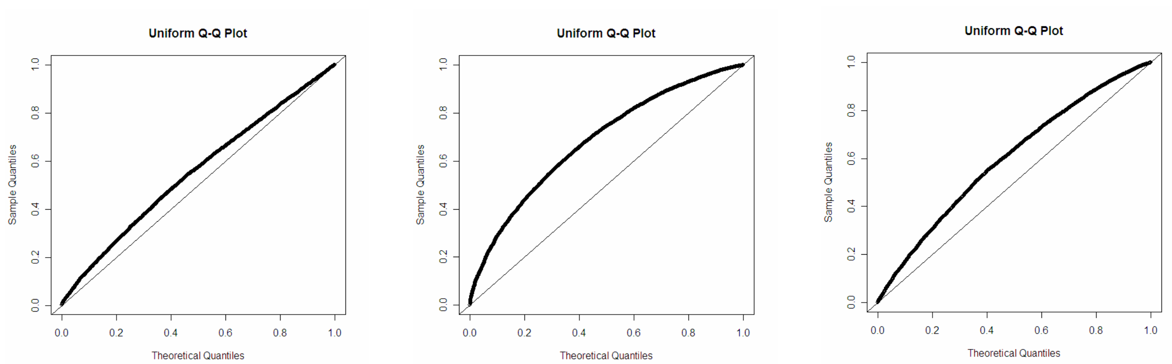

(2) $n_1 = n_2 = n_3 = 100$

**Figure A.4 Fiducial P-Values under $H_0$ (2.4)**

Conservative Test—P-value (3.9)

(A.4.1) $X_1 \underset{D}{=} X_2 \sim \mathrm{N}(0,1)$, $X_3 \sim \mathrm{N}(1.56767,1)$, $\underset{\sim}{l}^T = (-1/2,-1/2,1)$, 5000 Iterations (rep = 1000)

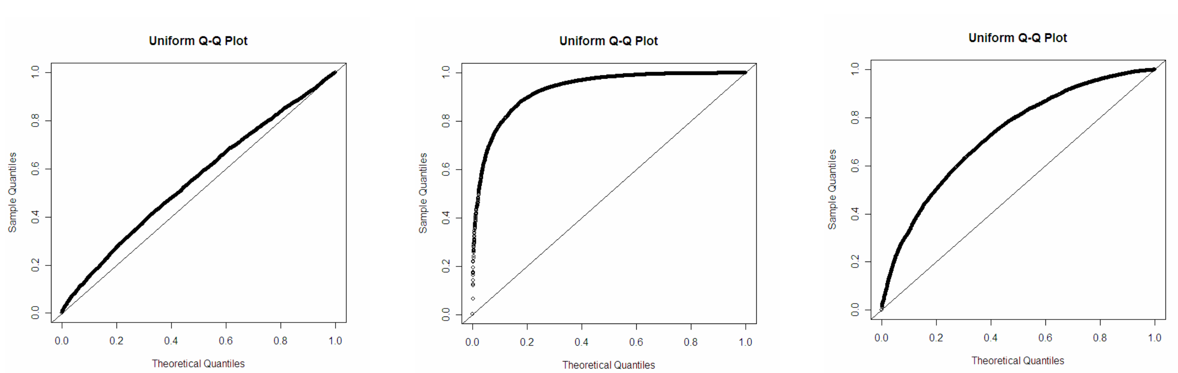
Figure A.4.1 Uniform Q-Q Plots of Fiducial P- Values (Conservative test)




(1) $n_1 = n_2 = n_3 = 10$ (2) $n_1 = 10$, $n_2 = 15$, $n_3 = 20$ (3) $n_1 = 100$, $n_2 = 110$, $n_3 = 120$

(A.4.2) $X_1 \sim \mathrm{N}(0,1)$, $X_2 \sim \mathrm{N}(0,(\sqrt{10})^2)$, $X_3 \sim \mathrm{N}(3,(\sqrt{2.74316})^2)$, $\underset{\sim}{l}^T = (-1/2,-1/2,1)$, B=5000 Iterations (rep = 1000)

Figure A.4.2 Uniform Q-Q Plots of Fiducial P- Values (Conservative test)




(1) $n_1 = n_2 = n_3 = 10$ (2) $n_1 = 10$, $n_2 = 50$, $n_3 = 20$ (3) $n_1 = 100$, $n_2 = 120$, $n_3 = 150$

**Figure A.5 Fiducial P-Values under $H_0$ (2.4)**

Replication Test—P-value (3.12)

(A.5.1) $X_1 \underset{D}{=} X_2 \sim \mathrm{N}(0,1)$, $X_3 \sim \mathrm{N}(1.56767,1)$, $\underline{l}^T = (-1/2,-1/2,1)$, 5000 Iterations (rep = 1000)

Figure A.5.1 The Uniform Q-Q Plot of Fiducial P- Value (Replication test)


(1) $n_1$=10, $n_2$=30, $n_3$=90 (2) $n_1$=100, $n_2$=110, $n_3$=120 (3) $n_1$=100, $n_2$=200, $n_3$=500

(A.5.2) $X_1 \sim \mathrm{N}(0,1)$, $X_2 \sim \mathrm{N}(0,(\sqrt{10})^2)$, $X_3 \sim \mathrm{N}(3,(\sqrt{2.74316})^2)$, $\underline{l}^T = (-1/2,-1/2,1)$, 5000 Iterations (rep = 1000)

Figure A.5.2 Uniform Q-Q Plots of Fiducial P- Value (Replication test)

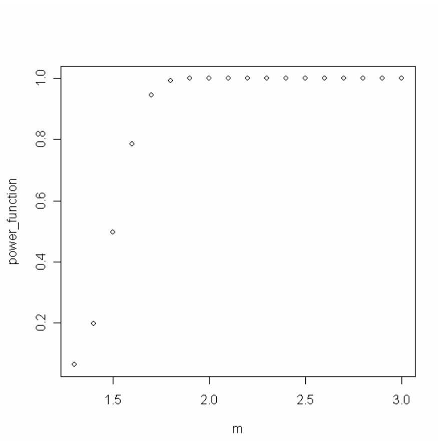

(1) n1=10, n2=200, n3=30 (2) n1=n2=n3=100 (3) n1=100, n2=500, n3=200

**Figure A.6 Fiducial P-Value -- Power ($H_0$ :(2.4) )**

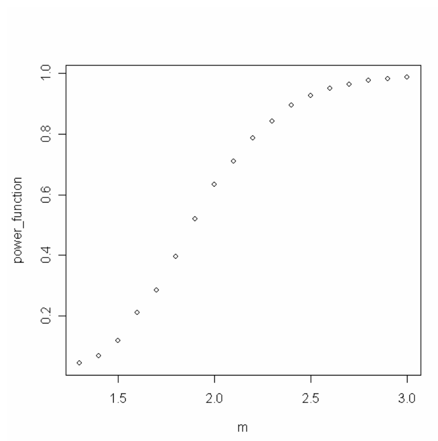
Replication Test (p-value (3.12))—Power

$X_1 \sim \mathrm{N}(0,1)$, $X_2 \sim \mathrm{N}(0,(\sqrt{10})^2)$, $X_3 \sim \mathrm{N}(5,(\sqrt{12.5087})^2)$, $\underset{\sim}{l}^T = (-1/2, -1/2, 1)$,
1000 Iterations (rep = 500) , m = $n_\pi$ .

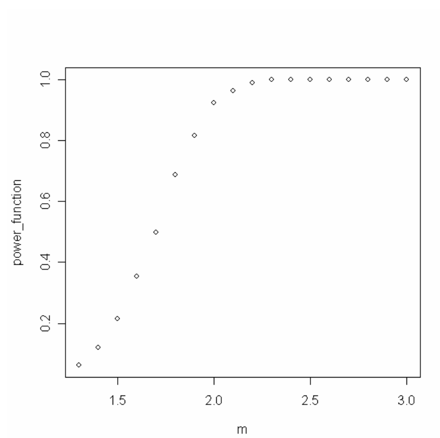
Figure A.6 Power vs $n_\pi$




(1) $n_1 = n_2 = n_3 = 100$




(2) $n_1 = 10$, $n_2 = 15$, $n_3 = 20$




(3) $n_1 = 30$, $n_2 = 40$, $n_3 = 30$

## APPENDIX B

**Table B.1: Estimated Type I Error Probabilities for LRT and PBL, $\pi = \pi_0$**

Table B.1.1 $K$=3, $\underline{l}^T = (-1/2, -1/2, 1)$, (a) $(\mu_2) = (0.5)$, $\alpha$ =*0.05*

| $\pi$ | small samples | | | medium samples | | | large samples | | |
|---|---|---|---|---|---|---|---|---|---|
| | *Method1* | | *Method2* | *Method1* | | *Method2* | *Method1* | | *Method2* |
| | *LRT* | *PBL* | *PBL* | *LRT* | *PBL* | *PBL* | *LRT* | *PBL* | *PBL* |
| *0.55* | 0.085 | 0.065 | 0.065 | 0.055 | 0.05 | 0.06 | 0.065 | 0.065 | 0.05 |
| *0.65* | 0.065 | 0.065 | 0.05 | 0.065 | 0.06 | 0.06 | 0.045 | 0.045 | 0.06 |
| *0.75* | 0.05 | 0.03 | 0.03 | 0.045 | 0.05 | 0.04 | 0.05 | 0.055 | 0.035 |
| *0.85* | 0.051 | 0.062 | 0.025 | 0.050 | 0.065 | 0.05 | 0.055 | 0.055 | 0.045 |

Table B.1.2 $K$=3, $\underline{l}^T = (-1/2, -1/2, 1)$, (b) $(\mu_2, \mu_3) = (0.5, 2)$, $\alpha$ =*0.05*

| $\pi$ | small samples | | | medium samples | | | large samples | | |
|---|---|---|---|---|---|---|---|---|---|
| | *Method1* | | *Method2* | *Method1* | | *Method2* | *Method1* | | *Method2* |
| | *LRT* | *PBL* | *PBL* | *LRT* | *PBL* | *PBL* | *LRT* | *PBL* | *PBL* |
| *0.55* | 0.07 | 0.06 | 0.05 | 0.04 | 0.035 | 0.045 | 0.065 | 0.055 | 0.045 |
| *0.65* | 0.055 | 0.05 | 0.055 | 0.055 | 0.04 | 0.06 | 0.07 | 0.055 | 0.03 |
| *0.75* | 0.085 | 0.055 | 0.04 | 0.05 | 0.04 | 0.06 | 0.065 | 0.065 | 0.04 |
| *0.85* | 0.041 | 0.052 | 0.021 | 0.047 | 0.052 | 0.026 | 0.04 | 0.045 | 0.055 |

Note:

(1) The gray color indicates that the entry is smaller than the lower bound of the approximate .95 confidence interval.
(2) The LRT p-value is constructed by using a chi-square distribution with df=1.

**Table B.2 Estimated Type I Error Probabilities for LRT and PBL, $\pi < \pi_0$**

Table B.2.1 Estimated type I error probabilities, $K$=3, $\underline{l}^T = (-1/2, -1/2, 1)$,
(a) $(\mu_2) = (0.5)$, $\alpha$ =0.05, $\pi_0 = 0.85$ (method1)

| $\pi$ | LRT | | | | PBL | | | |
|---|---|---|---|---|---|---|---|---|
| | *0.65* | *0.70* | *0.75* | *0.80* | *0.65* | *0.70* | *0.75* | *0.80* |
| *Small samples* | 0.575 | 0.37 | 0.18 | 0.09 | 0.575 | 0.395 | 0.21 | 0.11 |
| *Medium samples* | 0.835 | 0.715 | 0.335 | 0.105 | 0.84 | 0.69 | 0.33 | 0.11 |
| *Large samples* | 1 | 1 | 0.96 | 0.47 | 1 | 1 | 0.955 | 0.46 |

Table B.2.2 Estimated type I error probabilities, $K$=3, $\underline{l}^T = (-1/2, -1/2, 1)$,
(b) $(\mu_2, \mu_3) = (0.5, 2)$, $\pi_0 = 0.75$ (method1)

| $\pi$ | LRT | | | | PBL | | | |
|---|---|---|---|---|---|---|---|---|
| | *0.60* | *0.65* | *0.70* | *0.74* | *0.60* | *0.65* | *0.70* | *0.74* |
| *Small samples* | 0.28 | 0.2 | 0.095 | 0.04 | 0.285 | 0.18 | 0.09 | 0.035 |
| *Medium samples* | 0.425 | 0.215 | 0.1 | 0.065 | 0.415 | 0.215 | 0.095 | 0.05 |
| *Large samples* | 0.99 | 0.815 | 0.355 | 0.045 | 0.99 | 0.815 | 0.34 | 0.05 |

Table B.2.3 Estimated type I error probabilities for PBL, $K$=3, $\underline{l}^T = (-1/2, -1/2, 1)$,
$\pi < \pi_0$ (method2)

| $\pi$ | (a) $(\mu_2) = (0.5)$, $\pi_0 = 0.85$ | | | | (b) $(\mu_2, \mu_3) = (0.5, 2)$, $\pi_0 = 0.75$ | | | |
|---|---|---|---|---|---|---|---|---|
| | *0.65* | *0.70* | *0.75* | *0.80* | *0.60* | *0.65* | *0.70* | *0.74* |
| *Small samples* | 0 | 0 | 0.005 | 0.005 | 0.015 | 0.01 | 0.025 | 0.06 |
| *Medium samples* | 0 | 0 | 0 | 0 | 0 | 0.005 | 0.002 | 0.025 |
| *Large samples* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.003 |

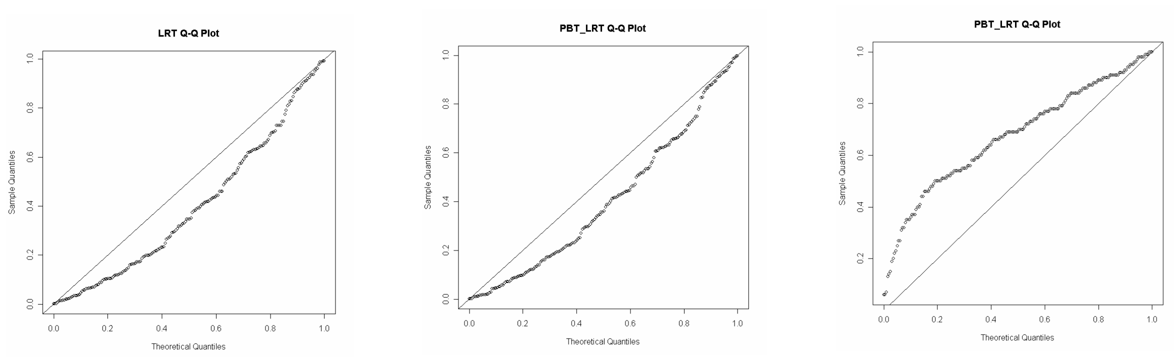
Note:

(1) The rose color indicates that the entry is greater than the upper bound of the approximate .95 confidence interval.

(2) The LRT p-value is constructed by using a chi-square distribution with df=1.

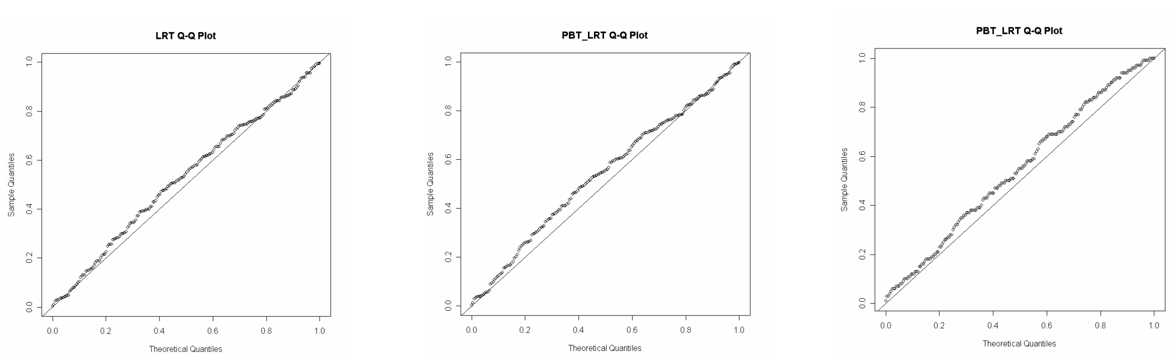
**Figure B.1 QQ Plots for Comparing Method1 and Method2**

Figure B.1.1 $K$=3, $\underline{l}^T = (-1/2, -1/2, 1)$, (a) $(\mu_2) = (0.5)$, $\alpha$ =0.05, $\pi_0 = 0.85$, $\pi$ =*0.80*, medium samples




(Method 1) (Method 1) (Method 2)

Figure B.1.2 $K$=3, $\underline{l}^T = (-1/2, -1/2, 1)$, (b) $(\mu_2, \mu_3) = (0.5, 2)$, $\alpha$ =0.05, $\pi_0 = 0.75$, $\pi$ =*0.74*), medium samples

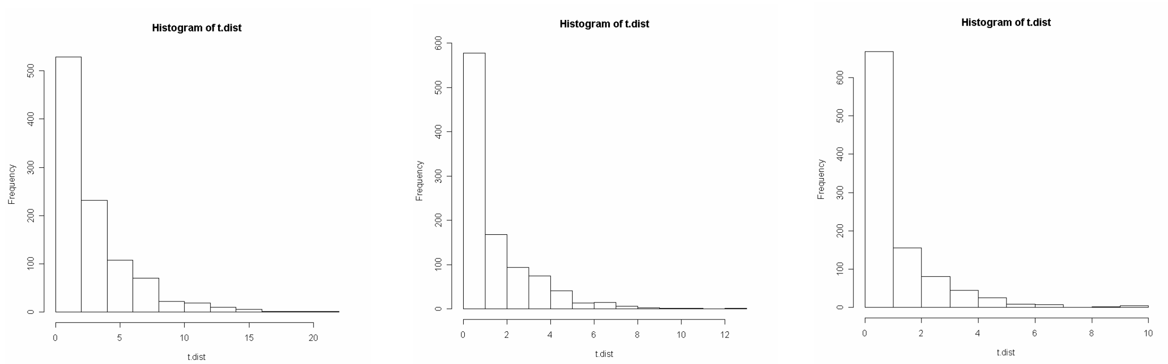



(Method 1) (Method 1) (Method 2)

**Figure B.2 Histograms of the Test Statistic ( $\lambda$ )**

$\pi_0 = 0.75$ , $K$=3, $\underline{l}^T = (-1/2, -1/2, 1)$ , (a) $(\mu_2) = (0.5)$, medium samples.
(generate 1000 data sets)

Figure B.2.1 Using Medthod 1 ( $\pi_0$=0.75 ) to obtain $\lambda$

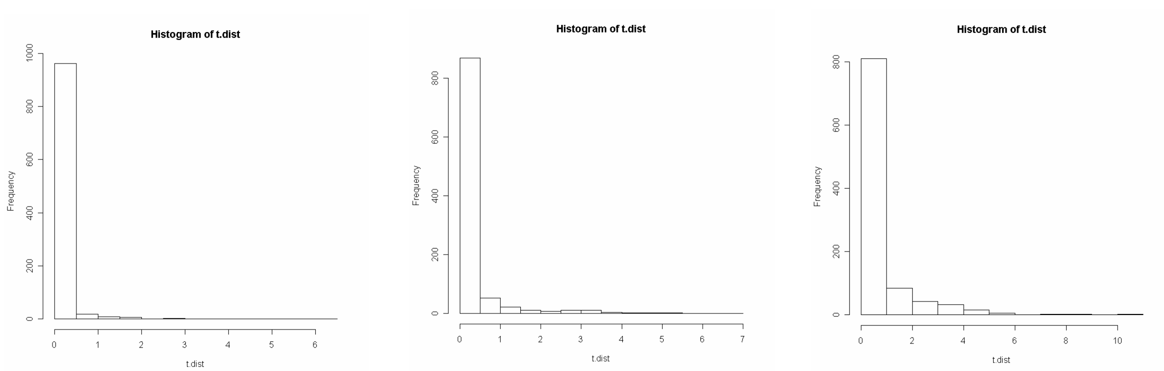



(a) $\pi^* = 0.65$
Mean: 2.736233
Variance: 8.996257

(b) $\pi^* = 0.70$
Mean: 1.401272
Variance: 3.095475

(c) $\pi^* = 0.75$
Mean: 1.051681
Variance: 2.098876

Figure B.2.2 Using Method 2 ( max( $L_j$ )=max( $\log L(\underline{\mu}, \underline{\sigma}^2)_j$ ) ) to obtain $\lambda$




(a) $\pi^* = 0.65$
Mean: 0.06337245
Variance: 0.1103547

(b) $\pi^* = 0.70$
Mean: 0.2563152
Variance: 0.5746209

(c) $\pi^* = 0.75$
Mean: 0.6334252
Variance: 1.740241

**Figure B.3 QQ Plots of the *PBL* P-Value for Different Gaps**

$K$=3, $\underline{l}^T = (-1/2, -1/2, 1)$, (b) $(\mu_2, \mu_3) = (0.5, 2)$, $\alpha$ =0.05, $\pi = \pi_0 = 0.75\,(PBL)$

Figure B.3.1 Small samples

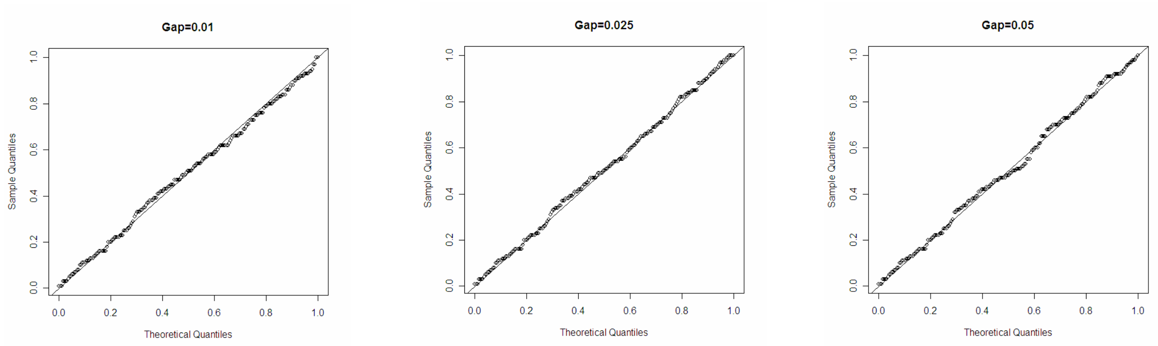



Figure B.3.2 Medium samples

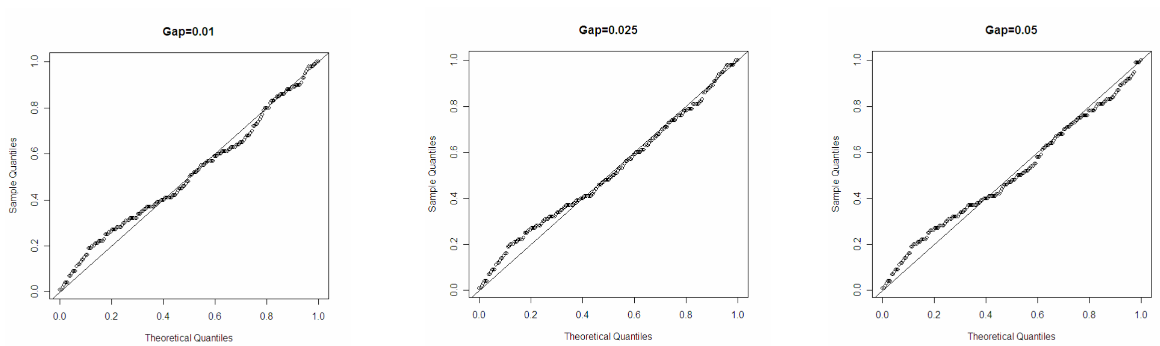
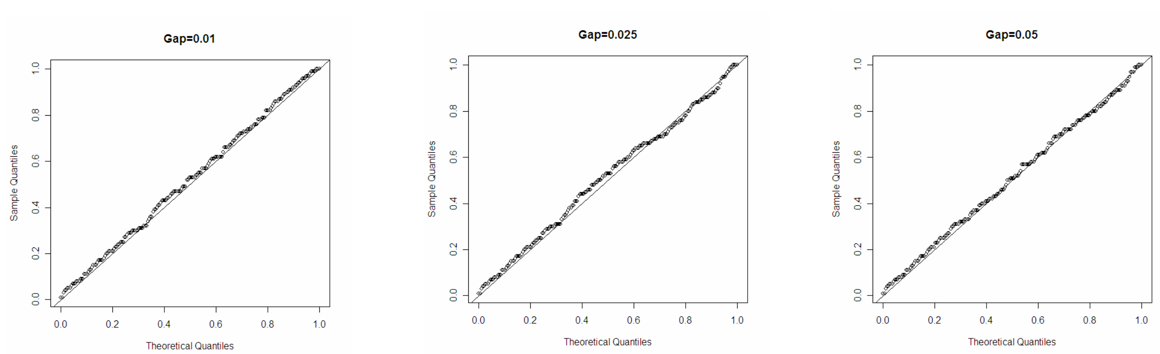



Figure B.3.3 Large samples



# APPENDIX C

**Table C.1 Estimated Type I Error Probabilities for $IS_{LIN}$ (F), $\pi = \pi_0$**

Table C.1.1 $K$=3, $\underline{l}^T = (-1/2, -1/2, 1)$, (a) $(\mu_2) = (0)$, $\alpha$ =0.05

| $\pi$ | small samples | | | | | medium samples | | | | | large samples | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *Plug* | *A_T* | *A_Z* | *PBp* | *PBL* | *Plug* | *A_T* | *A_Z* | *PBp* | *PBL* | *Plug* | *A_T* | *A_Z* | *PBp* | *PBL* |
| *0.55* | 0.055 | 0.035 | 0.035 | 0.055 | 0.055 | 0.05 | 0.05 | 0.05 | 0.05 | 0.055 | 0.06 | 0.06 | 0.06 | 0.055 | 0.03 |
| *0.65* | 0.055 | 0.035 | 0.03 | 0.05 | 0.045 | 0.055 | 0.05 | 0.05 | 0.05 | 0.075 | 0.06 | 0.06 | 0.06 | 0.045 | 0.04 |
| *0.75* | 0.065 | 0.045 | 0.045 | 0.06 | 0.055 | 0.05 | 0.05 | 0.05 | 0.055 | 0.055 | 0.06 | 0.055 | 0.06 | 0.055 | 0.05 |
| *0.85* | 0.04 | 0.025 | 0.025 | 0.035 | 0.031 | 0.04 | 0.04 | 0.04 | 0.04 | 0.075 | 0.05 | 0.045 | 0.045 | 0.06 | 0.085 |

Table C.1.2 $K$=3, $\underline{l}^T = (-1/2, -1/2, 1)$, (a) $(\mu_2) = (1)$, $\alpha$ =0.05

| $\pi$ | small samples | | | | | medium samples | | | | | large samples | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *Plug* | *A_T* | *A_Z* | *PBp* | *PBL* | *Plug* | *A_T* | *A_Z* | *PBp* | *PBL* | *Plug* | *A_T* | *A_Z* | *PBp* | *PBL* |
| *0.75* | 0.045 | 0.035 | 0.035 | 0.05 | 0.055 | 0.05 | 0.05 | 0.05 | 0.065 | 0.08 | 0.035 | 0.035 | 0.035 | 0.04 | 0.065 |
| *0.80* | 0.035 | 0.02 | 0.02 | 0.025 | 0.06 | 0.05 | 0.045 | 0.05 | 0.045 | 0.07 | 0.05 | 0.05 | 0.05 | 0.05 | 0.04 |
| *0.85* | 0.065 | 0.045 | 0.045 | 0.026 | 0.025 | 0.05 | 0.04 | 0.045 | 0.05 | 0.03 | 0.035 | 0.03 | 0.03 | 0.035 | 0.075 |
| *0.90* | 0.06 | 0.03 | 0.03 | 0.006 | 0.006 | 0.06 | 0.06 | 0.065 | 0.011 | 0.016 | 0.04 | 0.035 | 0.035 | 0.035 | 0.065 |

Table C.1.3 $K$=3, $\underline{l}^T = (-1/2, -1/2, 1)$, (b) $(\mu_2, \mu_3) = (0.5, 1)$, $\alpha$ =0.05

| $\pi$ | small samples | | | | | medium samples | | | | | large samples | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *Plug* | *A_T* | *A_Z* | *PBp* | *PBL* | *Plug* | *A_T* | *A_Z* | *PBp* | *PBL* | *Plug* | *A_T* | *A_Z* | *PBp* | *PBL* |
| *0.55* | 0.04 | 0.04 | 0.04 | 0.03 | 0.055 | 0.03 | 0.03 | 0.03 | 0.03 | 0.05 | 0.07 | 0.07 | 0.07 | 0.065 | 0.07 |
| *0.65* | 0.055 | 0.045 | 0.045 | 0.06 | 0.035 | 0.04 | 0.04 | 0.04 | 0.045 | 0.07 | 0.035 | 0.035 | 0.035 | 0.035 | 0.03 |
| *0.75* | 0.05 | 0.035 | 0.035 | 0.04 | 0.05 | 0.045 | 0.035 | 0.035 | 0.04 | 0.035 | 0.06 | 0.055 | 0.055 | 0.055 | 0.035 |
| *0.80* | 0.06 | 0.035 | 0.035 | 0.065 | 0.05 | 0.04 | 0.03 | 0.03 | 0.05 | 0.035 | 0.065 | 0.06 | 0.06 | 0.065 | 0.065 |

Table C.1.4 $K$=3, $\underline{l}^T = (-1/2, -1/2, 1)$, (b) $(\mu_2, \mu_3) = (0.5, 5)$, $\alpha$ =0.05

| $\pi$ | small samples | | | | | medium samples | | | | | large samples | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *Plug* | *A_T* | *A_Z* | *PBp* | *PBL* | *Plug* | *A_T* | *A_Z* | *PBp* | *PBL* | *Plug* | *A_T* | *A_Z* | *PBp* | *PBL* |
| *0.55* | 0.035 | 0.035 | 0.035 | 0.03 | 0.06 | 0.05 | 0.05 | 0.05 | 0.05 | 0.065 | 0.04 | 0.04 | 0.04 | 0.035 | 0.03 |
| *0.65* | 0.045 | 0.045 | 0.045 | 0.04 | 0.055 | 0.03 | 0.035 | 0.035 | 0.035 | 0.04 | 0.045 | 0.045 | 0.045 | 0.045 | 0.07 |
| *0.75* | 0.03 | 0.03 | 0.03 | 0.03 | 0.026 | 0.055 | 0.06 | 0.06 | 0.04 | 0.035 | 0.06 | 0.06 | 0.055 | 0.055 | 0.08 |
| *0.85* | 0.045 | 0.045 | 0.045 | 0 | 0 | 0.05 | 0.05 | 0.045 | 0 | 0 | 0.055 | 0.05 | 0.055 | 0.055 | 0.035 |

Table C.1.5 $K$=5, $\underset{\sim}{l}^T = (-1/3, -1/3, -1/3, 1/2, 1/2)$, (a) $(\mu_2, \mu_3, \mu_4) = (0, 0, 1)$, $\alpha$ =0.05

| $\pi$ | small samples | | | | | medium samples | | | | | large samples | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *Plug* | *A_T* | *A_Z* | *PBp* | *PBL* | *Plug* | *A_T* | *A_Z* | *PBp* | *PBL* | *Plug* | *A_T* | *A_Z* | *PBp* | *PBL* |
| *0.75* | 0.04 | 0.025 | 0.025 | 0.035 | 0.04 | 0.065 | 0.055 | 0.055 | 0.075 | 0.08 | 0.055 | 0.05 | 0.05 | 0.055 | 0.06 |
| *0.80* | 0.045 | 0.005 | 0.005 | 0.04 | 0.04 | 0.06 | 0.06 | 0.06 | 0.07 | 0.07 | 0.04 | 0.035 | 0.035 | 0.05 | 0.055 |
| *0.85* | 0.04 | 0.02 | 0.02 | 0.04 | 0.055 | 0.045 | 0.03 | 0.03 | 0.04 | 0.035 | 0.05 | 0.04 | 0.04 | 0.045 | 0.05 |
| *0.90* | 0.075 | 0.025 | 0.025 | 0.035 | 0.045 | 0.025 | 0.02 | 0.02 | 0.04 | 0.04 | 0.05 | 0.045 | 0.045 | 0.055 | 0.055 |

Table C.1.6 $K$=5, $\underset{\sim}{l}^T = (-1/3, -1/3, -1/3, 1/2, 1/2)$, (a) $(\mu_2, \mu_3, \mu_4) = (0.5, 0.5, 1)$, $\alpha$ =0.05

| $\pi$ | small samples | | | | | medium samples | | | | | large samples | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *Plug* | *A_T* | *A_Z* | *PBp* | *PBL* | *Plug* | *A_T* | *A_Z* | *PBp* | *PBL* | *Plug* | *A_T* | *A_Z* | *PBp* | *PBL* |
| *0.75* | 0.06 | 0.035 | 0.035 | 0.075 | 0.075 | 0.055 | 0.045 | 0.045 | 0.05 | 0.05 | 0.04 | 0.035 | 0.035 | 0.055 | 0.055 |
| *0.80* | 0.02 | 0.005 | 0.005 | 0.03 | 0.045 | 0.05 | 0.04 | 0.04 | 0.05 | 0.055 | 0.055 | 0.045 | 0.045 | 0.035 | 0.04 |
| *0.85* | 0.055 | 0.045 | 0.04 | 0.051 | 0.051 | 0.055 | 0.03 | 0.03 | 0.055 | 0.05 | 0.05 | 0.045 | 0.045 | 0.05 | 0.05 |
| *0.90* | 0.065 | 0.02 | 0.025 | 0.045 | 0.045 | 0.045 | 0.04 | 0.04 | 0.045 | 0.045 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 |

Table C.1.7 $K$=5, $\underset{\sim}{l}^T = (-1/3, -1/3, -1/3, 1/2, 1/2)$, (b) $(\mu_2, \mu_3, \mu_4, \mu_5) = (0, 0, 2, 2)$, $\alpha$ =0.05

| $\pi$ | small samples | | | | | medium samples | | | | | large samples | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *Plug* | *A_T* | *A_Z* | *PBp* | *PBL* | *Plug* | *A_T* | *A_Z* | *PBp* | *PBL* | *Plug* | *A_T* | *A_Z* | *PBp* | *PBL* |
| *0.75* | 0.06 | 0.05 | 0.05 | 0.045 | 0.045 | 0.04 | 0.04 | 0.04 | 0.04 | 0.04 | 0.05 | 0.05 | 0.05 | 0.045 | 0.045 |
| *0.80* | 0.05 | 0.04 | 0.04 | 0.04 | 0.04 | 0.05 | 0.05 | 0.05 | 0.055 | 0.055 | 0.04 | 0.03 | 0.03 | 0.05 | 0.05 |
| *0.85* | 0.04 | 0.025 | 0.025 | 0.020 | 0.025 | 0.045 | 0.04 | 0.04 | 0.03 | 0.03 | 0.045 | 0.04 | 0.04 | 0.055 | 0.055 |
| *0.90* | 0.06 | 0.015 | 0.015 | 0.027 | 0.011 | 0.04 | 0.035 | 0.035 | 0.02 | 0.015 | 0.075 | 0.065 | 0.065 | 0.06 | 0.065 |

Table C.1.8 $K$=5, $\underset{\sim}{l}^T = (-1/3, -1/3, -1/3, 1/2, 1/2)$, (b) $(\mu_2, \mu_3, \mu_4, \mu_5) = (0.5, 1, 3, 3.5)$, $\alpha$ =0.05

| $\pi$ | small samples | | | | | medium samples | | | | | large samples | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *Plug* | *A_T* | *A_Z* | *PBp* | *PBL* | *Plug* | *A_T* | *A_Z* | *PBp* | *PBL* | *Plug* | *A_T* | *A_Z* | *PBp* | *PBL* |
| *0.75* | 0.04 | 0.03 | 0.03 | 0.025 | 0.02 | 0.035 | 0.035 | 0.035 | 0.025 | 0.025 | 0.045 | 0.045 | 0.045 | 0.04 | 0.04 |
| *0.80* | 0.06 | 0.05 | 0.05 | 0.036 | 0.036 | 0.08 | 0.08 | 0.08 | 0.061 | 0.061 | 0.065 | 0.055 | 0.05 | 0.065 | 0.06 |
| *0.85* | 0.04 | 0.025 | 0.025 | 0.005 | 0.005 | 0.04 | 0.035 | 0.035 | 0.026 | 0.031 | 0.04 | 0.035 | 0.035 | 0.035 | 0.03 |
| *0.90* | 0.045 | 0.035 | 0.035 | 0.022 | 0.005 | 0.05 | 0.05 | 0.05 | 0.021 | 0.015 | 0.04 | 0.03 | 0.03 | 0.03 | 0.035 |

Table C.1.9 $K$=7, (a) $(\mu_2, \mu_3, \mu_4, \mu_5, \mu_6) = (0.5, 0.5, 0.5, 1, 1)$, $\alpha$ =0.05,

$\underset{\sim}{l}^T = (-1/6, -1/6, -1/6, -1/6, -1/6, -1/6, 1)$

| $\pi$ | small samples | | | | | medium samples | | | | | large samples | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *Plug* | *A_T* | *A_Z* | *PBp* | *PBL* | *Plug* | *A_T* | *A_Z* | *PBp* | *PBL* | *Plug* | *A_T* | *A_Z* | *PBp* | *PBL* |
| *0.70* | 0.055 | 0.04 | 0.04 | 0.055 | 0.055 | 0.06 | 0.05 | 0.045 | 0.045 | 0.05 | 0.04 | 0.04 | 0.04 | 0.05 | 0.05 |
| *0.75* | 0.09 | 0.05 | 0.05 | 0.06 | 0.06 | 0.055 | 0.055 | 0.055 | 0.055 | 0.055 | 0.055 | 0.055 | 0.055 | 0.045 | 0.045 |
| *0.80* | 0.095 | 0.065 | 0.065 | 0.051 | 0.051 | 0.075 | 0.065 | 0.065 | 0.065 | 0.065 | 0.08 | 0.08 | 0.08 | 0.075 | 0.075 |
| *0.85* | 0.065 | 0.045 | 0.045 | 0.011 | 0.006 | 0.065 | 0.045 | 0.045 | 0.051 | 0.056 | 0.08 | 0.08 | 0.08 | 0.07 | 0.07 |

Table C.1.10 $K$=7, (a) $(\mu_2, \mu_3, \mu_4, \mu_5, \mu_6) = (1, 1, 1, 1, 1)$, $\alpha$ =0.05,

$\underset{\sim}{l}^T = (-1/6, -1/6, -1/6, -1/6, -1/6, -1/6, 1)$

| $\pi$ | small samples | | | | | medium samples | | | | | large samples | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *Plug* | *A_T* | *A_Z* | *PBp* | *PBL* | *Plug* | *A_T* | *A_Z* | *PBp* | *PBL* | *Plug* | *A_T* | *A_Z* | *PBp* | *PBL* |
| *0.70* | 0.07 | 0.055 | 0.055 | 0.065 | 0.065 | 0.06 | 0.055 | 0.055 | 0.065 | 0.065 | 0.045 | 0.04 | 0.04 | 0.045 | 0.045 |
| *0.75* | 0.08 | 0.05 | 0.05 | 0.07 | 0.065 | 0.065 | 0.055 | 0.055 | 0.055 | 0.055 | 0.06 | 0.06 | 0.06 | 0.06 | 0.065 |
| *0.80* | 0.06 | 0.04 | 0.04 | 0.035 | 0.035 | 0.045 | 0.045 | 0.045 | 0.06 | 0.06 | 0.035 | 0.035 | 0.035 | 0.045 | 0.045 |
| *0.85* | 0.075 | 0.04 | 0.04 | 0.006 | 0.006 | 0.045 | 0.03 | 0.03 | 0.035 | 0.035 | 0.065 | 0.065 | 0.06 | 0.05 | 0.04 |

Table C.1.11 $K$=7, (b) $(\mu_2, \mu_3, \mu_4, \mu_5, \mu_6, \mu_7) = (0, 0.5, 0.5, 0.5, 0.5, 2)$, $\alpha$ =0.05

$\underset{\sim}{l}^T = (-1/6, -1/6, -1/6, -1/6, -1/6, -1/6, 1)$

| $\pi$ | small samples | | | | | medium samples | | | | | large samples | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *Plug* | *A_T* | *A_Z* | *PBp* | *PBL* | *Plug* | *A_T* | *A_Z* | *PBp* | *PBL* | *Plug* | *A_T* | *A_Z* | *PBp* | *PBL* |
| *0.75* | 0.045 | 0.045 | 0.045 | 0.035 | 0.035 | 0.045 | 0.04 | 0.035 | 0.045 | 0.045 | 0.06 | 0.06 | 0.06 | 0.05 | 0.05 |
| *0.80* | 0.055 | 0.05 | 0.055 | 0.011 | 0.011 | 0.07 | 0.065 | 0.065 | 0.06 | 0.06 | 0.065 | 0.055 | 0.055 | 0.06 | 0.06 |
| *0.85* | 0.06 | 0.02 | 0.025 | 0.02 | 0 | 0.045 | 0.04 | 0.04 | 0.011 | 0.011 | 0.055 | 0.05 | 0.055 | 0.05 | 0.05 |
| *0.90* | 0.05 | 0.035 | 0.035 | 0 | 0 | 0.05 | 0.045 | 0.045 | 0 | 0 | 0.03 | 0.035 | 0.035 | 0.035 | 0.035 |

Table C.1.12 $K$=7, (b) $(\mu_2, \mu_3, \mu_4, \mu_5, \mu_6, \mu_7) = (0.5, 1, 1, 1, 1.5, 3)$, $\alpha$ =0.05,

$\underset{\sim}{l}^T = (-1/6, -1/6, -1/6, -1/6, -1/6, -1/6, 1)$

| $\pi$ | small samples | | | | | medium samples | | | | | large samples | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *Plug* | *A_T* | *A_Z* | *PBp* | *PBL* | *Plug* | *A_T* | *A_Z* | *PBp* | *PBL* | *Plug* | *A_T* | *A_Z* | *PBp* | *PBL* |
| *0.75* | 0.055 | 0.05 | 0.05 | 0.03 | 0.03 | 0.06 | 0.06 | 0.06 | 0.06 | 0.06 | 0.06 | 0.06 | 0.06 | 0.045 | 0.045 |
| *0.80* | 0.055 | 0.035 | 0.035 | 0.011 | 0.011 | 0.065 | 0.06 | 0.06 | 0.065 | 0.06 | 0.06 | 0.055 | 0.06 | 0.06 | 0.06 |
| *0.85* | 0.09 | 0.075 | 0.075 | 0 | 0 | 0.065 | 0.055 | 0.05 | 0 | 0 | 0.055 | 0.055 | 0.055 | 0.05 | 0.05 |
| *0.90* | 0.09 | 0.07 | 0.07 | 0 | 0 | 0.055 | 0.05 | 0.05 | 0 | 0 | 0.055 | 0.055 | 0.055 | 0.010 | 0.010 |

**Table C.2 Estimated Type I Error Probabilities of P-Values for IS$_{LIN}$(F), $\pi < \pi_0$**

Table C.2.1 $K$=3, $\underline{l}^T = (-1/2, -1/2, 1)$, (a) $(\mu_2) = (0.5)$, $\alpha$ =0.05, $\pi_0 = 0.85$

| $\pi$ | small samples | | | | | medium samples | | | | | large samples | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *Plug* | *A_T* | *A_Z* | *PBp* | *PBL* | *Plug* | *A_T* | *A_Z* | *PBp* | *PBL* | *Plug* | *A_T* | *A_Z* | *PBp* | *PBL* |
| *0.65* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *0.75* | 0.01 | 0.01 | 0.01 | 0.01 | 0.005 | 0.005 | 0.005 | 0.005 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *0.80* | 0.015 | 0.01 | 0.01 | 0.005 | 0.005 | 0.01 | 0.01 | 0.01 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *0.84* | 0.025 | 0.025 | 0.025 | 0.01 | 0.01 | 0.035 | 0.035 | 0.035 | 0.030 | 0.030 | 0.04 | 0.04 | 0.04 | 0.04 | 0.045 |

Table C.2.2 $K$=3, $\underline{l}^T = (-1/2, -1/2, 1)$, (b) $(\mu_2, \mu_3) = (0.5, 2)$, $\alpha$ =0.05, $\pi_0 = 0.75$

| $\pi$ | small samples | | | | | medium samples | | | | | large samples | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *Plug* | *A_T* | *A_Z* | *PBp* | *PBL* | *Plug* | *A_T* | *A_Z* | *PBp* | *PBL* | *Plug* | *A_T* | *A_Z* | *PBp* | *PBL* |
| *0.60* | 0.005 | 0.005 | 0.005 | 0.005 | 0.015 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *0.65* | 0.01 | 0.01 | 0.01 | 0.005 | 0.01 | 0.005 | 0.005 | 0.005 | 0.005 | 0.005 | 0 | 0 | 0 | 0 | 0 |
| *0.70* | 0.03 | 0.025 | 0.025 | 0.025 | 0.025 | 0.015 | 0.015 | 0.015 | 0.02 | 0.02 | 0 | 0 | 0 | 0 | 0 |
| *0.74* | 0.07 | 0.055 | 0.055 | 0.055 | 0.060 | 0.035 | 0.035 | 0.035 | 0.025 | 0.025 | 0.04 | 0.04 | 0.04 | 0.03 | 0.03 |

Table C.2.3 $K$=5, $\underline{l}^T = (-1/3, -1/3, -1/3, 1/2, 1/2)$, $\alpha$ =0.05, $\pi_0 = 0.85$

(a) $(\mu_2, \mu_3, \mu_4) = (0.5, 0.5, 1)$

| $\pi$ | small samples | | | | | medium samples | | | | | large samples | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *Plug* | *A_T* | *A_Z* | *PBp* | *PBL* | *Plug* | *A_T* | *A_Z* | *PBp* | *PBL* | *Plug* | *A_T* | *A_Z* | *PBp* | *PBL* |
| *0.65* | 0 | 0 | 0 | 0 | 0.005 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *0.75* | 0 | 0 | 0 | 0.005 | 0.005 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *0.80* | 0.01 | 0 | 0 | 0.005 | 0.015 | 0 | 0 | 0 | 0.005 | 0.005 | 0 | 0 | 0 | 0 | 0 |
| *0.84* | 0.045 | 0.01 | 0.01 | 0.045 | 0.040 | 0.025 | 0.025 | 0.02 | 0.03 | 0.035 | 0.01 | 0.01 | 0.005 | 0.01 | 0.01 |

Table C.2.4 $K$=7, $\underline{l}^T = (-1/6, -1/6, -1/6, -1/6, -1/6, -1/6, 1)$, $\alpha$ =0.05, $\pi_0 = 0.75$,

(a) $(\mu_2, \mu_3, \mu_4, \mu_5, \mu_6) = (0.5, 0.5, 0.5, 1, 1)$

| $\pi$ | small samples | | | | | medium samples | | | | | large samples | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *Plug* | *A_T* | *A_Z* | *PBp* | *PBL* | *Plug* | *A_T* | *A_Z* | *PBp* | *PBL* | *Plug* | *A_T* | *A_Z* | *PBp* | *PBL* |
| *0.60* | 0 | 0 | 0 | 0 | 0.005 | 0 | 0 | 0 | 0 | 0.005 | 0 | 0 | 0 | 0 | 0 |
| *0.65* | 0.015 | 0.015 | 0.015 | 0.015 | 0.015 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *0.70* | 0.02 | 0.02 | 0.015 | 0.035 | 0.035 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0 | 0 | 0 | 0 | 0 |
| *0.74* | 0.06 | 0.04 | 0.04 | 0.035 | 0.04 | 0.045 | 0.03 | 0.035 | 0.04 | 0.04 | 0.025 | 0.025 | 0.025 | 0.045 | 0.045 |

**Table C.3 Cochran's Test**

$H_0$: The powers are equally effective

$H_a$: There is a difference in effectiveness among Powers

Table C.3.1. $K$=3, $\underline{l}^T = (-1/2, -1/2, 1)$, (a) $(\mu_2) = (0.5)$, $\pi = 0.65$, $\alpha = 0.05$

| | $\pi_{H_a}$=0.66 | $\pi_{H_a}$=0.76 | $\pi_{H_a}$=0.80 | $\pi_{H_a}$=0.88 | $\pi_{H_a}$=0.98 |
|---|---|---|---|---|---|
| Small samples | Do not reject | Reject | Reject | Reject | Do not reject |
| Medium samples | Do not reject | Do not reject | Do not reject | Do not reject | Do not reject |
| Large samples | Do not reject | Do not reject | Do not reject | Do not reject | Do not reject |

Table C.3.2. $K$=3, $\underline{l}^T = (-1/2, -1/2, 1)$, (b) $(\mu_2, \mu_3) = (0.5, 2)$, $\pi = 0.75$, $\alpha = 0.05$

| | $\pi_{H_a}$=0.76 | $\pi_{H_a}$=0.82 | $\pi_{H_a}$=0.85 | $\pi_{H_a}$=0.88 | $\pi_{H_a}$=0.95 |
|---|---|---|---|---|---|
| Small samples | Do not reject | Reject | Reject | Reject | Reject |
| Medium samples | Do not reject | Do not reject | Do not reject | Reject (p-value=0.0435) | Do not reject |
| Large samples | Do not reject | Do not reject | Do not reject | Do not reject | Do not reject |

Table C.3.3. $K$=5, $\underline{l}^T = (-1/3, -1/3, -1/3, 1/2, 1/2)$, (a) $(\mu_2, \mu_3, \mu_4) = (0.5, 0.5, 1)$, $\pi = 0.75$, $\alpha = 0.05$

| | $\pi_{H_a}$=0.76 | $\pi_{H_a}$=0.82 | $\pi_{H_a}$=0.85 | $\pi_{H_a}$=0.88 | $\pi_{H_a}$=0.95 |
|---|---|---|---|---|---|
| Small samples | Reject | Reject | Reject | Reject | Rejct |
| Medium samples | Reject | Reject | Reject | Reject | Do not reject |
| Large samples | Do not reject | Do not reject | Do not reject | Do not reject | Do not reject |

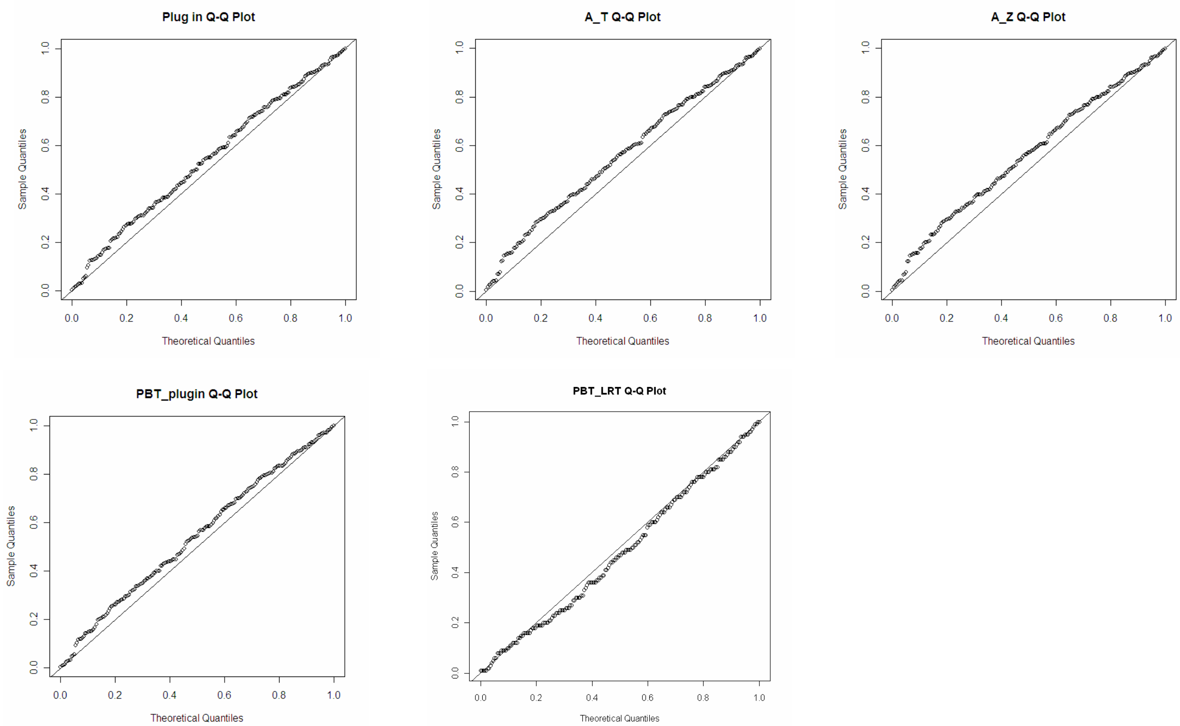
Table C.3.4. $K$=7, (a) $(\mu_2, \mu_3, \mu_4, \mu_5, \mu_6) = (0.5, 0.5, 0.5, 1, 1)$, $\pi = 0.70$, $\alpha = 0.05$, $\underline{l}^T = (-1/6, -1/6, -1/6, -1/6, -1/6, -1/6, 1)$

| | $\pi_{H_a}$=0.75 | $\pi_{H_a}$=0.82 | $\pi_{H_a}$=0.85 | $\pi_{H_a}$=0.88 | $\pi_{H_a}$=0.95 |
|---|---|---|---|---|---|
| Small samples | Reject | Reject | Do not reject | Reject | Do not reject |
| Medium samples | Do not reject | Do not reject | Do not reject | Do not reject | Do not reject |
| Large samples | Do not reject | Do not reject | Do not reject | Do not reject | Do not reject |

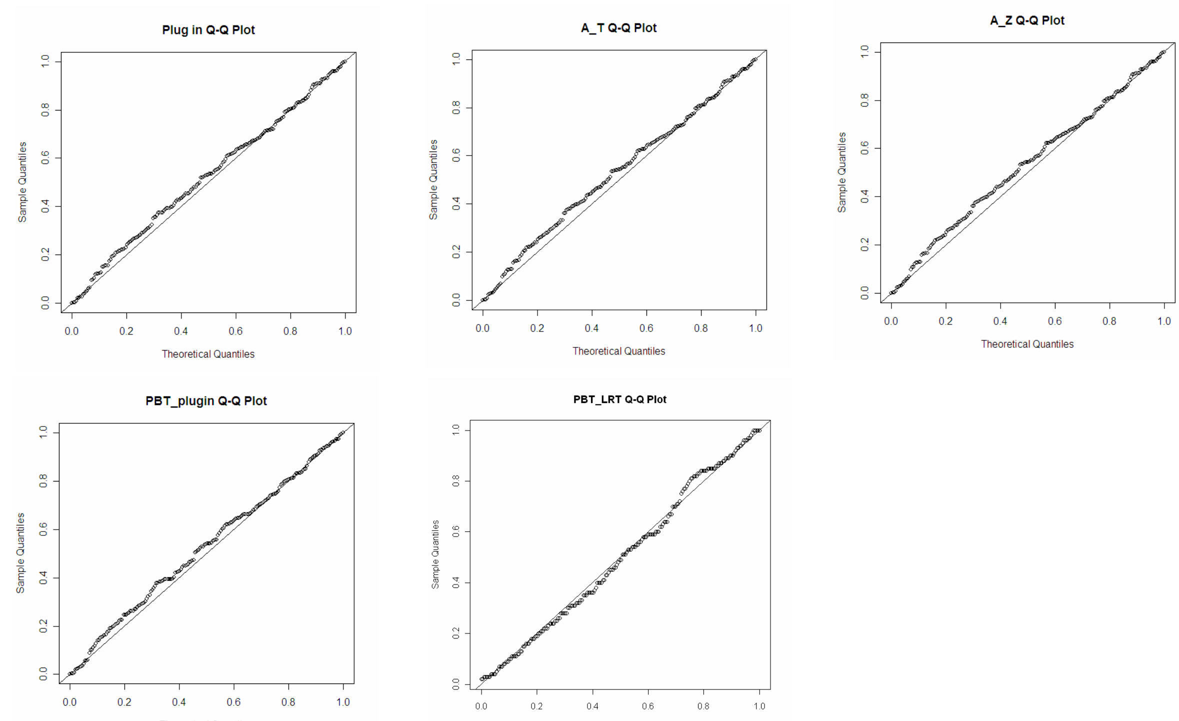
**Figure C.1 QQ Plots of the P-Values for $\text{IS}_{\text{LIN}}$ (F), $\pi = \pi_0$**

Figure C.1.1 $K$=3, $\underline{l}^T = (-1/2, -1/2, 1)$, (a) $(\mu_2) = (0.5)$, $\pi = 0.65$

(a) Small samples (10,15,12)




(b) Medium samples (30,50,20)



(c) Large samples (100,150,120)

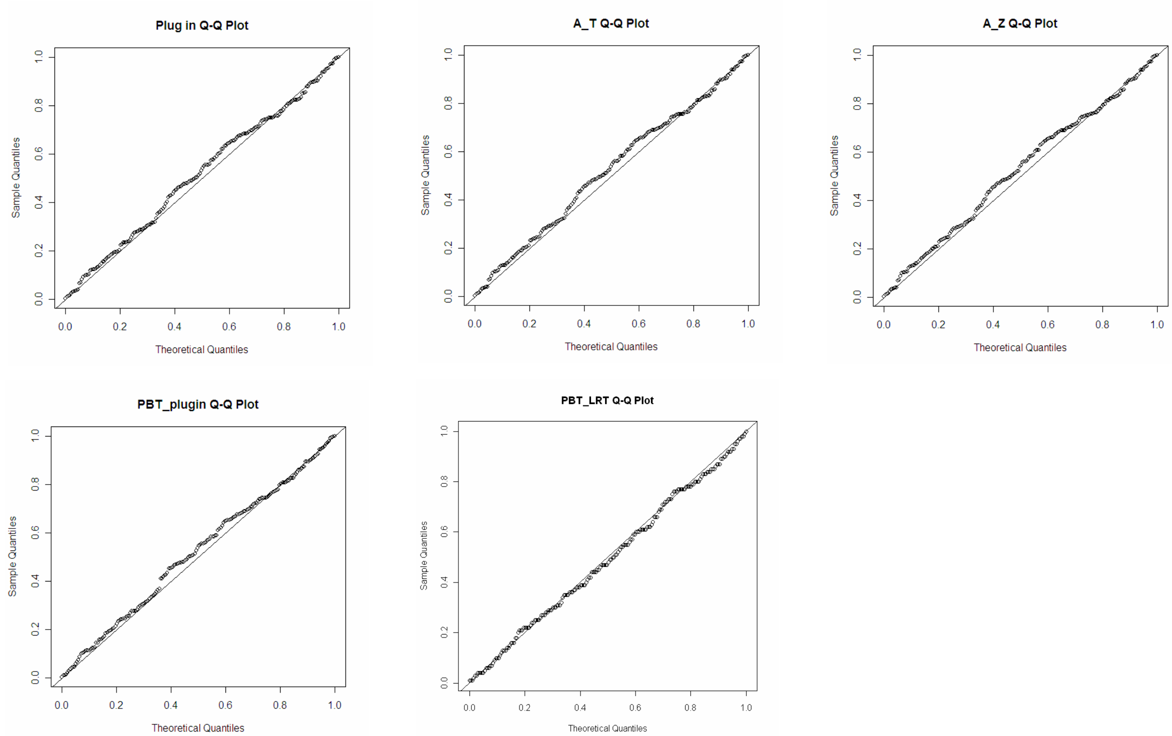
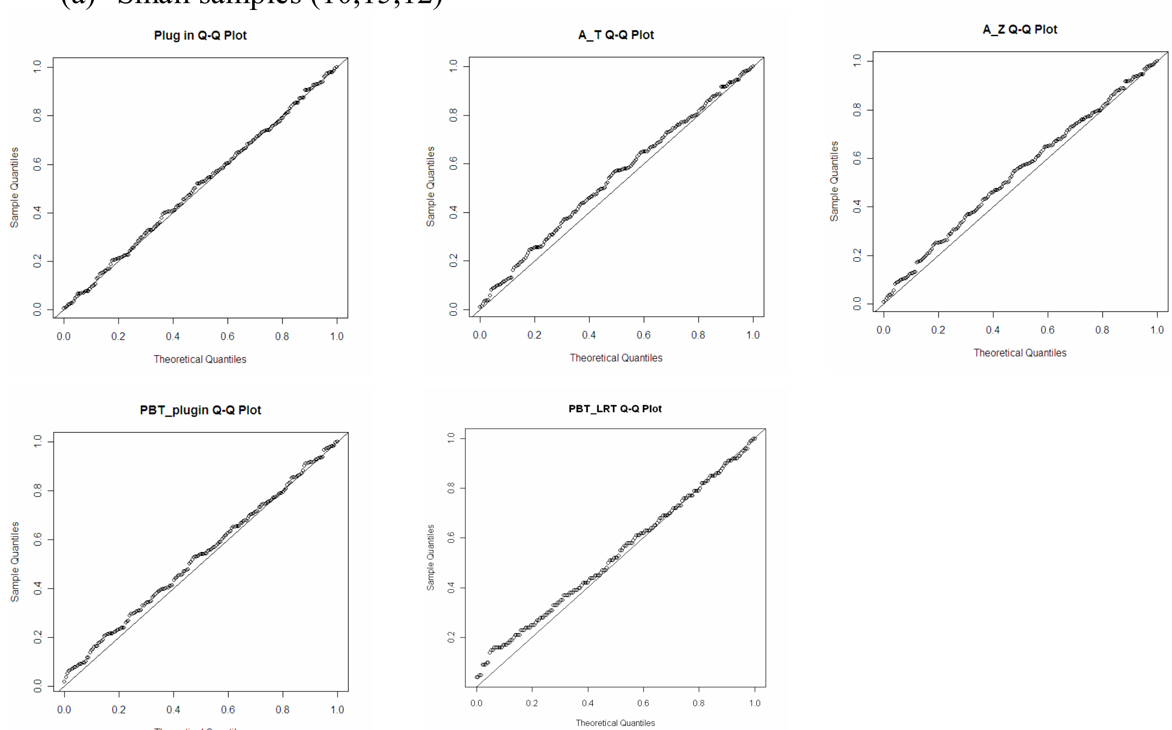
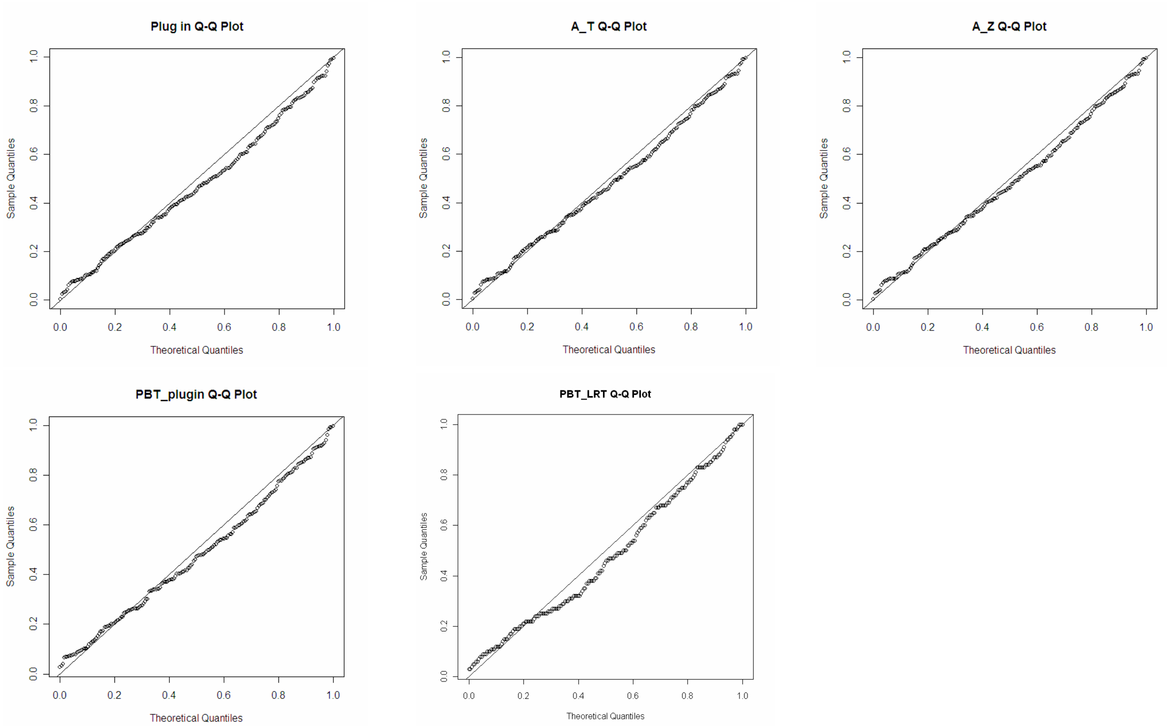



Figure C.1.2 $K$=3, $\underline{l}^T = (-1/2,-1/2,1)$, (b) $(\mu_2,\mu_3)=(0.5,\ 2)$, $\pi = 0.85$

(a) Small samples (10,15,12)



(b) Medium samples (30,50,20)




(c) Large samples (100,150,120)

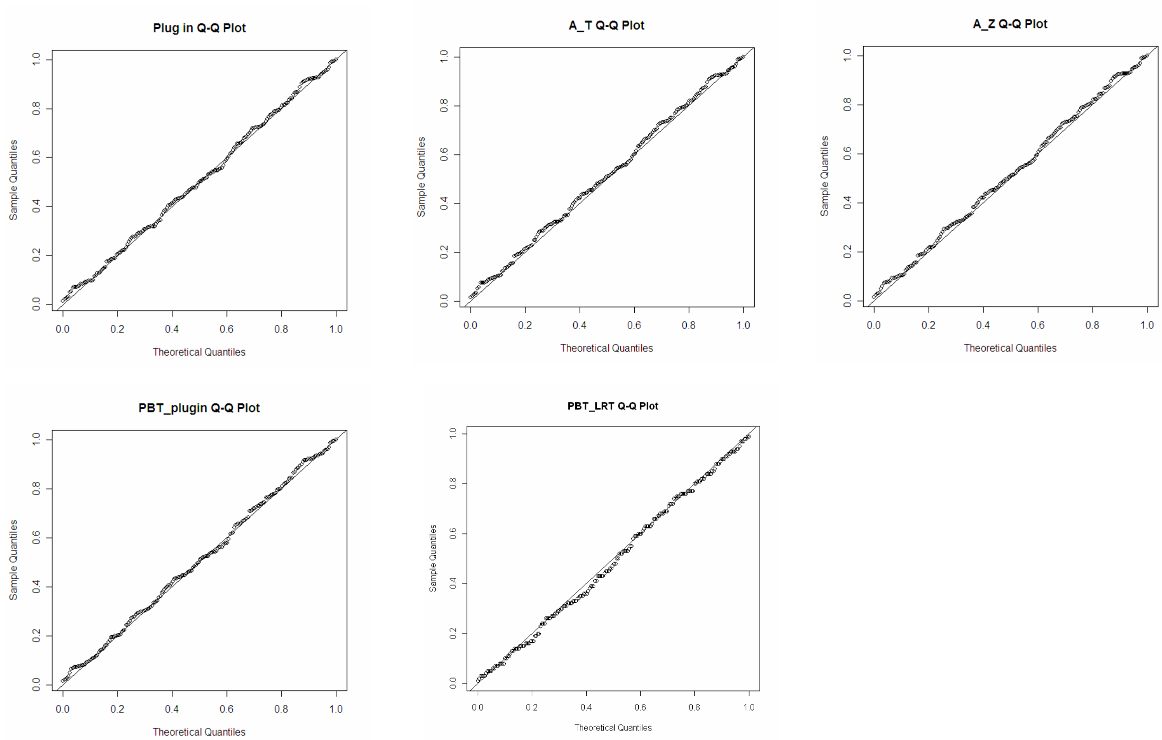


**Figure C.2 QQ Plots of the P-Values for $\mathrm{IS}_{\mathrm{LIN}}$ (F), $\pi < \pi_0$**, Medium Samples

Figure C.2.1 $K$=3, $\underline{l}^T = (-1/2, -1/2, 1)$, (a) $(\mu_2) = (0.5)$, $\alpha$ =0.05, $\pi_0 = 0.85$, $\pi$ =0.80.

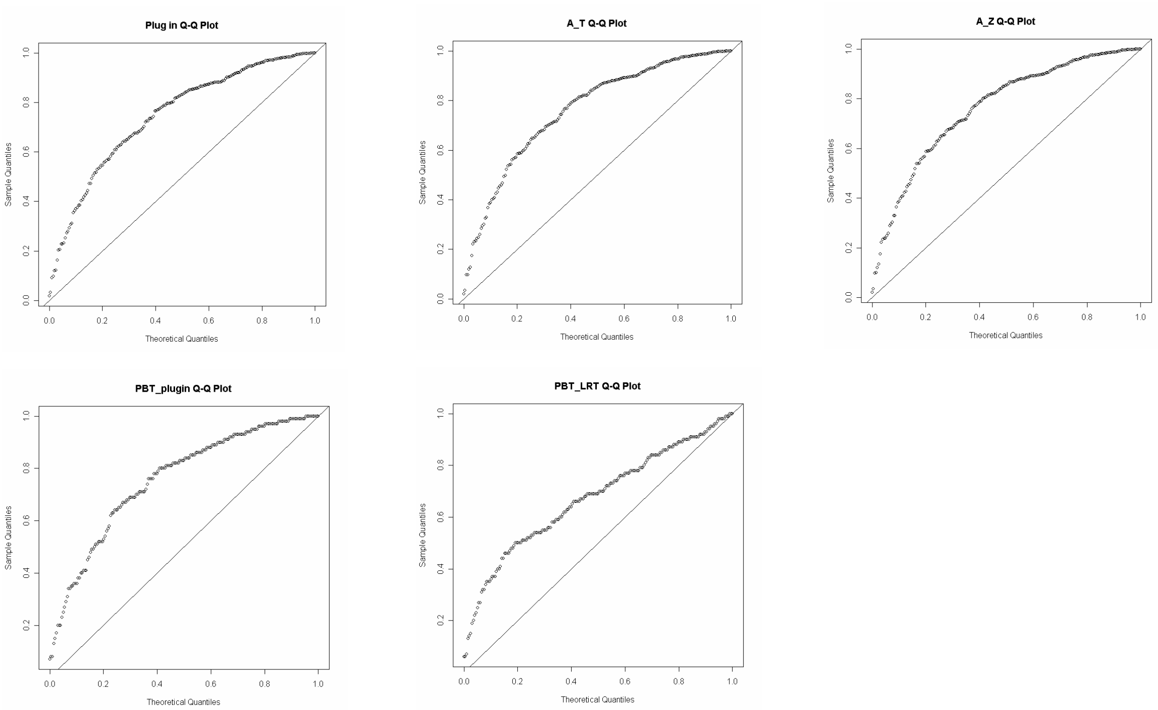



Figure C.2.2 $K$=3, $\underline{l}^T = (-1/2, -1/2, 1)$, (b) $(\mu_2, \mu_3) = (0.5, 2)$, $\alpha$ =0.05, $\pi_0 = 0.75$, $\pi$ =0.74.

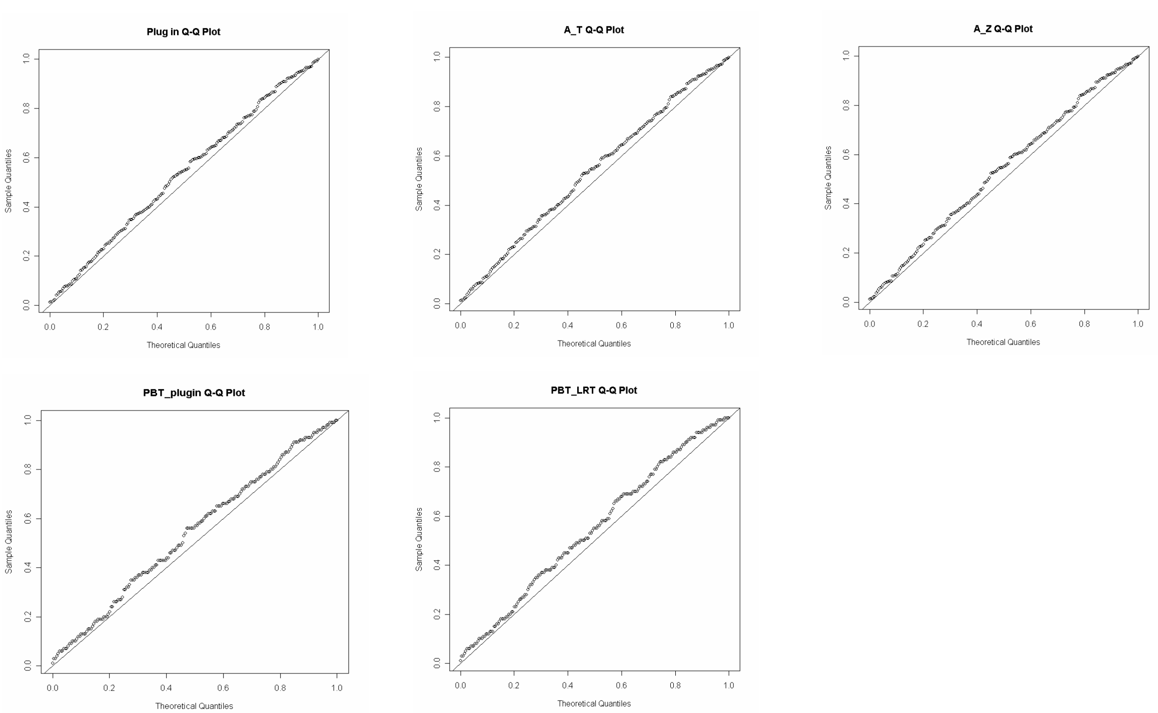
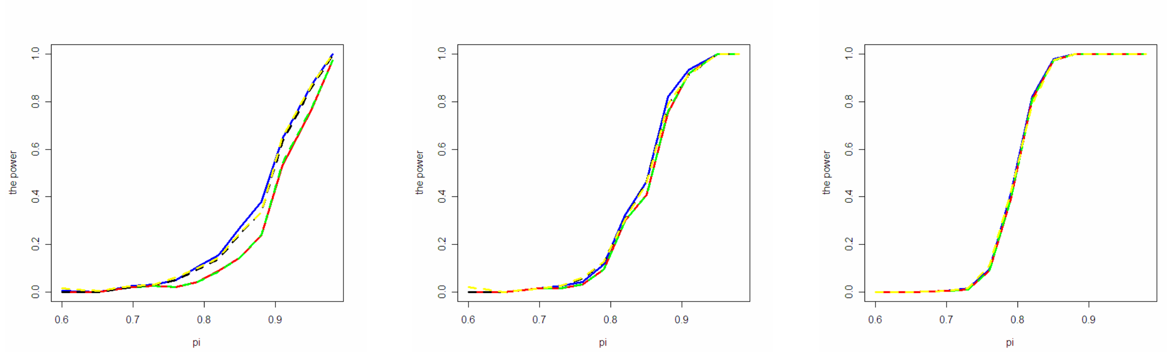


**Figure C.3 Power Simulation Results of Tests for $IS_{LIN}(\underline{F})$**

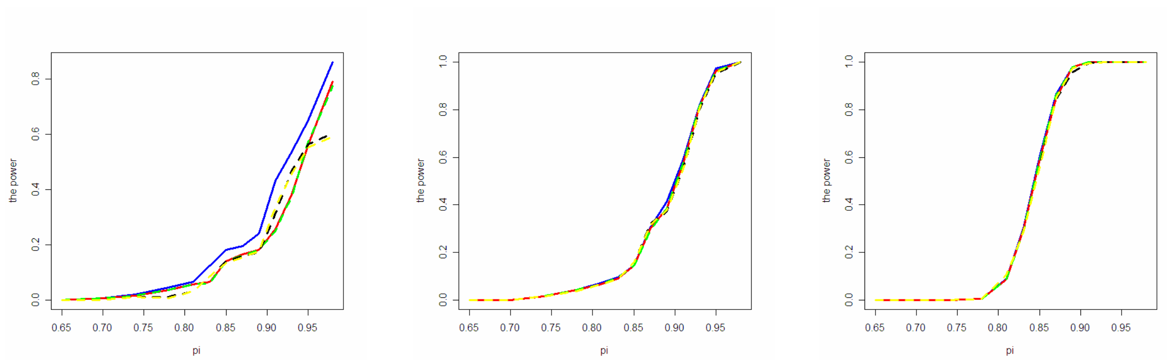
Figure C.3.1 $K$=5, $\underline{l}^T = (-1/3, -1/3, -1/3, 1/2, 1/2)$, (a) $(\mu_2, \mu_3, \mu_4)$ = (0.5, 0.5, 1), $\pi = 0.75$, $\alpha$ =0.05.




(a) small samples (b) medium samples (c) large samples

Blue: *Plug_in*
Red : *Average_T*
Green : *Average_Z*
Black: *PBT_plugin*
Yellow: *PBT_LRT*

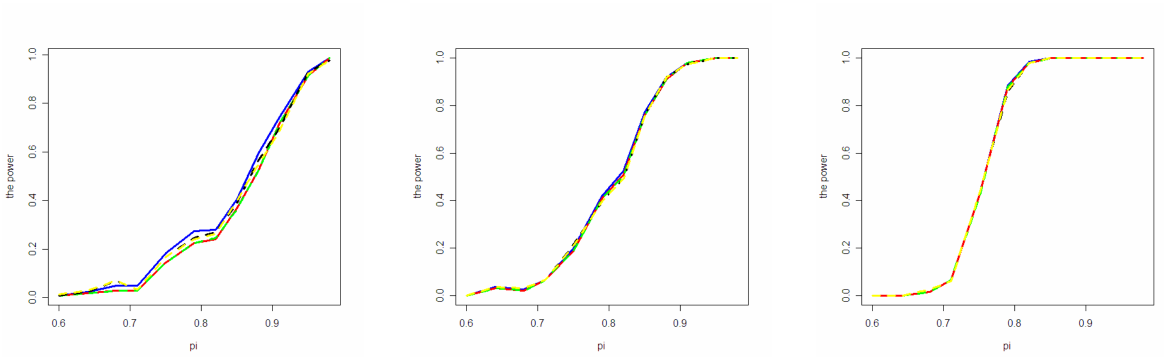
Figure C.3.2 $K$=5, $\underline{l}^T = (-1/3, -1/3, -1/3, 1/2, 1/2)$, (b) $(\mu_2, \mu_3, \mu_4, \mu_5) = (0.5, 1, 3, 3.5)$, $\pi = 0.80$ $\alpha$ =0.05.




(a) small samples (b) medium samples (c) large samples

Blue: *Plug_in*
Red : *Average_T*
Green : *Average_Z*
Black: *PBT_plugin*
Yellow: *PBT_LRT*

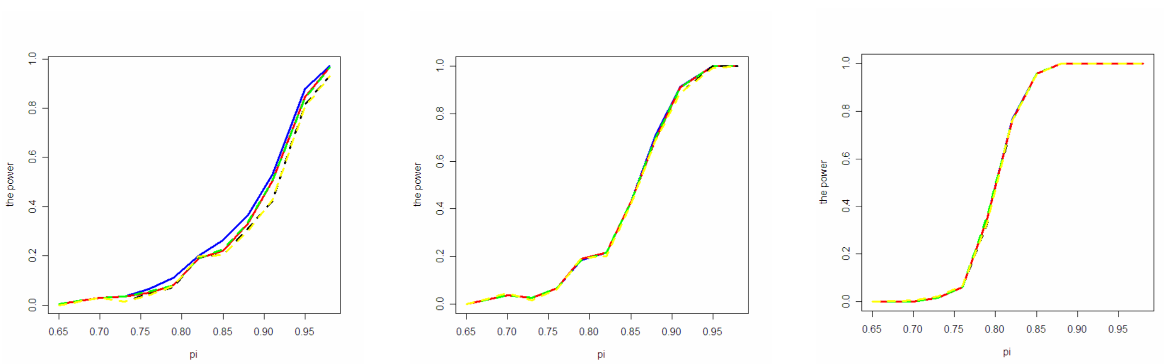
Figure C.3.3 $K$=7, (a) $(\mu_2, \mu_3, \mu_4, \mu_5, \mu_6)$ = (0.5, 0.5, 0.5, 1, 1), $\pi = 0.70$, $\alpha$ =0.05,
$\underset{\sim}{l}^T = (-1/6, -1/6, -1/6, -1/6, -1/6, -1/6, 1)$.




(a) small samples (b) medium samples (c) large samples

| | | | |
|---|---|---|---|
| Blue: | *Plug_in* | Black: | *PBT_plugin* |
| Red : | *Average_T* | Yellow: | *PBT_LRT* |
| Green : | *Average_Z* | | |

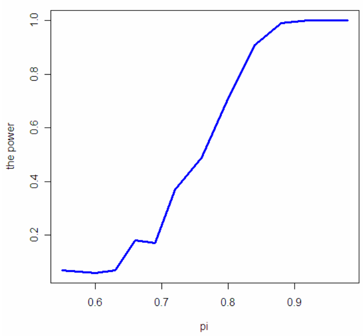
Figure C.3.4 $K$=7, (b) $(\mu_2, \mu_3, \mu_4, \mu_5, \mu_6, \mu_7)$ = (0.5, 1, 1, 1, 1.5, 3), $\pi = 0.75$ $\alpha$ =0.05
$\underset{\sim}{l}^T = (-1/6, -1/6, -1/6, -1/6, -1/6, -1/6, 1)$.



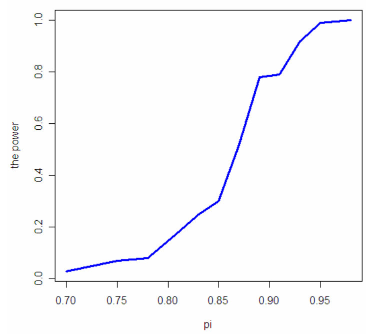

(a) small samples (b) medium samples (c) large samples

| | | | |
|---|---|---|---|
| Blue: | *Plug_in* | Black: | *PBT_plugin* |
| Red : | *Average_T* | Yellow: | *PBT_LRT* |
| Green : | *Average_Z* | | |

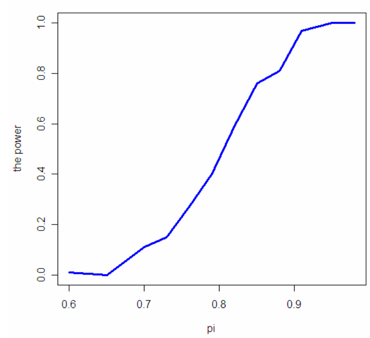
**Figure C.4 Power Simulation Results for $IS_{AV}(\underline{F})$**



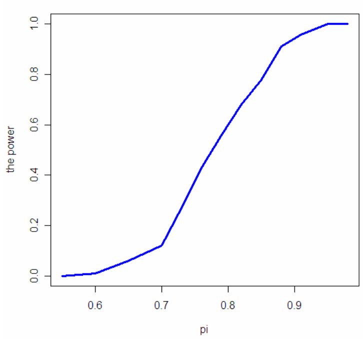

(C.4.1)



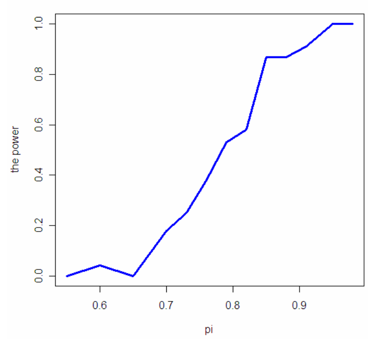

(C.4.2)



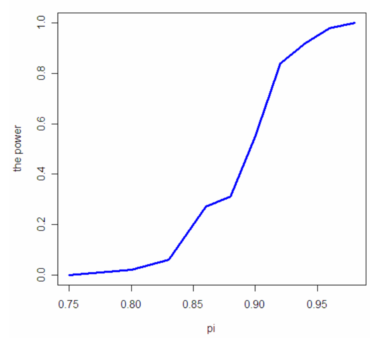

(C.4.3)



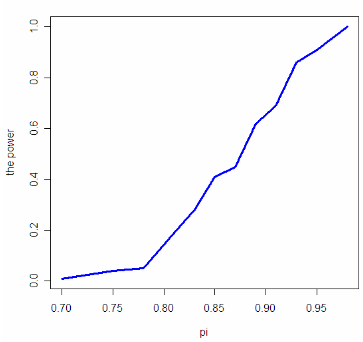

(C.4.4)



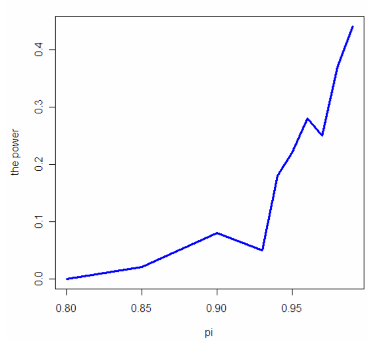

(C.4.5)




(C.4.6)




(C.4.7)




(C.4.8)

(C.4.1) $K$=3, (a) $(\mu_2) = (0.5)$ $\pi = 0.60$, n=(20,25,28),
(C.4.2) $K$=3, (a) $(\mu_2) = (1)$ $\pi = 0.70$, n=(20,25,28),
(C.4.3) $K$=3, (b) $(\mu_2, \mu_3)$ = (0.5, 1), $\pi = 0.65$, n=(20,25,28),
(C.4.4) $K$=3, (b) $(\mu_2, \mu_3)$ = (0.5, 2), $\pi = 0.65$, n=(20,25,28),
(C.4.5) $K$=3, (b) $(\mu_2, \mu_3)$ = (0.5, 5), $\pi = 0.65$, n=(20,25,28),
(C.4.6) $K$=5, (a) $(\mu_2, \mu_3, \mu_4)$ = (0,0,1) $\pi = 0.70$, n=(50,60,55,54,58),
(C.4.7) $K$=5, (a) $(\mu_2, \mu_3, \mu_4)$ = (0.5,0.5,1) $\pi = 0.65$, n=(20,25,22,24,26),
(C.4.8) $K$=5, (b) $(\mu_2, \mu_3, \mu_4, \mu_5)$ = (0.5, 1, 3, 3.5), $\pi = 0.77$, n=(20,25,22,24,26).

# APPENDIX D

## Result 1 R Code for Estimated Type I Error Probabilities for $IS_{LIN}$ (F) ($K$=3), Method 2

```
###############################################################################
# This Splus/R Function to compare the estimated P-Values of my Ph.D research
# ni = sample size for random variable xi
# Normality
# Ho: P(l'x>0)<=Pi       #K=3
#######################################################################################
#Obtain the MLE's

mle.iteration<-function(n1,n2,n3,x1,x2,x3,npai,gap){
m<-0        # counts iterations
diff<-1
l<-c(l1,l2,l3)
n<-c(n1,n2,n3)

sample.mean<-c(mean(x1),mean(x2),mean(x3))
sample.var<-c(var(x1),var(x2),var(x3))
mu.hat<-sample.mean
var.hat<-rep(0,3)

y<-rep(0,3)
z<-rep(0,3)

while( diff>gap) {
m<-m+1

var.hat[1]<-sum((x1-mu.hat[1])^2)/(n[1]*(1-l[1]*(npai^2)*(sample.mean[1]-mu.hat[1])/sum(l*mu.hat)))
var.hat[2]<-sum((x2-mu.hat[2])^2)/(n[2]*(1-l[2]*(npai^2)*(sample.mean[2]-mu.hat[2])/sum(l*mu.hat)))
var.hat[3]<-sum((x3-mu.hat[3])^2)/(n[3]*(1-l[3]*(npai^2)*(sample.mean[3]-mu.hat[3])/sum(l*mu.hat)))

if (is.na(var.hat[1])==TRUE) {
   results<-NA
   break
}

if (min(var.hat)<0) {
   #cat("var.hat<0","\n")
   return (NULL)
   break
}

  diff.var<-abs(z-var.hat)
  z<-var.hat

for (j in 1:3) {
y[j]<-sample.mean[j]+(l[j]*var.hat[j]*(npai*sqrt(sum(l^2*var.hat))-
sum(l*sample.mean)))/(n[j]*sum(l^2*var.hat/n))
}
  diff.mu<-abs(y-mu.hat)
  diff<-max(c(diff.mu,diff.var))
```

```
  mu.hat<-y

if (m>=800) {
   #cat("iteration more than 800 ",  "\n")
   return (NULL)
   break
}
results<-cbind(mu.hat,var.hat)
}
return (results)
}

##################################################################################
# function to get the test statistic T_star_plugin

Tstar_plug<-function(n1,n2,n3,x1,x2,x3){
a1<-l1^2*var(x1)/n1+l2^2*var(x2)/n2+l3^2*var(x3)/n3
T_p_plugin<-(l1*mean(x1)+l2*mean(x2)+l3*mean(x3))/sqrt(a1)
T_p_plugin
}

#########################################################
# function to get the test statistic T_star_LRT

Tstar_LRT1<-function(n1,n2,n3,x1,x2,x3,npai){
mle<-mle.iteration(n1,n2,n3,x1,x2,x3,npai,1e-5)

if (is.matrix(mle)==FALSE) {
   lambda<-NA
} else  {
S1.2<-var(x1)*(n1-1)/n1
S2.2<-var(x2)*(n2-1)/n2
S3.2<-var(x3)*(n3-1)/n3

lambda1<-(-2)*((n1/2)*log(S1.2/mle[1,2])+(n1/2)-sum((x1-mle[1,1])^2/(2*mle[1,2])))
lambda2<-(-2)*((n2/2)*log(S2.2/mle[2,2])+(n2/2)-sum((x2-mle[2,1])^2/(2*mle[2,2])))
lambda3<-(-2)*((n3/2)*log(S3.2/mle[3,2])+(n3/2)-sum((x3-mle[3,1])^2/(2*mle[3,2])))

lambda=lambda1+lambda2+lambda3
}
lambda
}

######################################################################################
max.log_mle<-function(n1,n2,n3,x1,x2,x3,npai){

mle<-mle.iteration(n1,n2,n3,x1,x2,x3,npai,1e-5)
if (is.matrix(mle)==FALSE) {
   log_mle<-NA
} else {
mle1<-(-1)*(n1/2)*log(2*(3.141593)*mle[1,2])-sum((x1-mle[1,1])^2/(2*mle[1,2]))
mle2<-(-1)*(n2/2)*log(2*(3.141593)*mle[2,2])-sum((x2-mle[2,1])^2/(2*mle[2,2]))
mle3<-(-1)*(n3/2)*log(2*(3.141593)*mle[3,2])-sum((x3-mle[3,1])^2/(2*mle[3,2]))
log_mle=mle1+mle2+mle3
}
```

```
result<-c(npai, log_mle)
result
}

###############################################################
# function to get the test statistic T_star_LRT

Tstar_LRT<-function(n1,n2,n3,x1,x2,x3){
try<-array(rep(0,length(pi_j)*2),c(length(pi_j),2))
 for (j in 1:length(pi_j))
{
 try[j,]<-max.log_mle(n1,n2,n3,x1,x2,x3,npai_j[j])
}
if (sum(is.na(try))>0) {
 try1<-try[-which(is.na(try[,2])),]
} else {
 try1<-try
}
  npai_max<-try1[which(try1[,2]==max(try1[,2]))]

  p_max<-Tstar_LRT1(n1,n2,n3,x1,x2,x3,npai_max)
  p_max
}

######################################################################################
general.p_value<-function(n1,n2,n3){

  x1<-rnorm(n1,mu1,sigma1)     # generate n Normal (0,1) random variables
  x2<-rnorm(n2,mu2,sigma2)      # generate n Normal (0,sqrt(10)) random variables
  x3<-rnorm(n3,mu3,sigma3)      # generate n Normal (3,sigma3) random variables

  var1_hat<-var(x1)*(n1-1)/rchisq(num,df=n1-1)
  var2_hat<-var(x2)*(n2-1)/rchisq(num,df=n2-1)
  var3_hat<-var(x3)*(n3-1)/rchisq(num,df=n3-1)

##################
# plug in p_value

a1<-l1^2*var(x1)/n1+l2^2*var(x2)/n2+l3^2*var(x3)/n3
T_p_plugin<-(l1*mean(x1)+l2*mean(x2)+l3*mean(x3))/sqrt(a1)
dem1<-(l1^2*var(x1)/n1)^2/(n1-1)+(l2^2*var(x2)/n2)^2/(n2-1)+(l3^2*var(x3)/n3)^2/(n3-1)
df1<-a1^2/dem1

delta1<-sqrt(l1^2*var(x1)+l2^2*var(x2)+l3^2*var(x3))*qnorm(pi)/sqrt(a1)
p_value.plugin<-1-pt(T_p_plugin, df=df1, ncp=delta1)

##################
# average p_value for T

a2<-l1^2*var1_hat/n1+l2^2*var2_hat/n2+l3^2*var3_hat/n3
b2<-((n1-1)*var(x1)/var1_hat+(n2-1)*var(x2)/var2_hat+(n3-1)*var(x3)/var3_hat)/(n1+n2+n3-3)
T_p_hat<-((l1*mean(x1)+l2*mean(x2)+l3*mean(x3))/sqrt(a2))/sqrt(b2)

delta2<-sqrt(l1^2*var1_hat+l2^2*var2_hat+l3^2*var3_hat)*qnorm(pi)/sqrt(a2)
p_value.T<-1-pt(T_p_hat, df=n1+n2+n3-3, ncp=delta2)
```

```
p_value.average_T<-mean(p_value.T)

#################
# average p_value for Z

Z_p_hat<-(l1*mean(x1)+l2*mean(x2)+l3*mean(x3)-
qnorm(pi)*sqrt(l1^2*var1_hat+l2^2*var2_hat+l3^2*var3_hat))/sqrt(a2)
p_value.Z<-1-pnorm(Z_p_hat)
p_value.average_Z<-mean(p_value.Z)

###############################################
# Parametric Bootstrap Test for T_plugin and LRT

T_p1<-Tstar_plug(n1,n2,n3,x1,x2,x3)
T_p2<-Tstar_LRT(n1,n2,n3,x1,x2,x3)

m<-99
mle2<-mle.iteration(n1,n2,n3,x1,x2,x3,qnorm(pi),1e-5)

if (is.matrix(mle2)==FALSE) {
   p_PBT_plug<-NA
   p_PBT_LRT<-NA

} else {
MLE.mu<-mle2[,1]
MLE.sigma<-sqrt(mle2[,2])

x1.boot1<-replicate(m, rnorm(n1, MLE.mu[1], MLE.sigma[1]))
x2.boot1<-replicate(m, rnorm(n2, MLE.mu[2], MLE.sigma[2]))
x3.boot1<-replicate(m, rnorm(n3, MLE.mu[3], MLE.sigma[3]))

# compute test statistics for each resample

T_boot.1<-rep(0,m)
T_boot.2<-rep(0,m)
for (j in 1:m) {
  T_boot.1[j]<-Tstar_plug(n1,n2,n3,x1.boot1[,j],x2.boot1[,j],x3.boot1[,j])
  T_boot.2[j]<-Tstar_LRT(n1,n2,n3,x1.boot1[,j],x2.boot1[,j],x3.boot1[,j])
}
reject1<-ifelse(T_boot.1>=T_p1,1,0)
reject.2<-ifelse(T_boot.2>=T_p2,1,0)
reject2<-reject.2[!is.na(reject.2)]
p_PBT_plug<-(sum(reject1)+1)/(m+1)
p_PBT_LRT<-(sum(reject2)+1)/(length(reject2)+1)
}

#################
# compute the p_value
p_value<-c(p_value.plugin, p_value.average_T,p_value.average_Z, p_PBT_plug, p_PBT_LRT)
p_value
}
##################################
set.seed(6543267)
num<-1000  # get 1000 number of independant var_hat from the chi-square distribution
B<-200  # get 100 data sets to do simulation
n1=30
```

```
n2=50
n3=20
l1=-0.5
l2=-0.5
l3=1
pi<-0.85
pi_j<-seq(0.50,pi,by=0.025)
npai_j<-qnorm(pi_j)
pi_true<-0.75

######################################################
#(a)equal variance
mu1<-0
mu2<-0.5
sigma1<-1
sigma2<-1
sigma3<-1
mu3<-sqrt(sigma3^2+(sigma1^2+sigma2^2)/4)*qnorm(pi_true)+(mu1+mu2)/2

#######################################################
#(b)sigmasq_i=mu_i+1
#mu1<-0
#sigma1<-1
#mu2<-0.5
#sigma2<-sqrt(mu2+1)
#mu3<-1
#sigma3<-sqrt(((mu3-(mu1+mu2)/2)/qnorm(pi_true))^2-(sigma1^2+sigma2^2)/4)

###############################################################
p.dist<-array(rep(0,5*B),c(5,B))
reject.05<-array(rep(0,5*B),c(5,B))
for (i in 1:B){
 p.dist[,i]<-general.p_value(n1,n2,n3)
reject.05[,i]<-ifelse(p.dist[,i]<=0.05,1,0)
}
####################################################################################
reject.05<-ifelse(p.dist<=0.05,1,0)
re1<-reject.05[1,]
re2<-reject.05[2,]
re3<-reject.05[3,]
re.4<-reject.05[4,]
re4<-re.4[!is.na(re.4)]
re.5<-reject.05[5,]
re5<-re.5[!is.na(re.5)]

typeI_error1.05<-sum(re1)/length(re1)
typeI_error2.05<-sum(re2)/length(re2)
typeI_error3.05<-sum(re3)/length(re3)
typeI_error4.05<-sum(re4)/length(re4)
typeI_error5.05<-sum(re5)/length(re5)

cat('TypeI error(0.05) plugin =', typeI_error1.05, '\n')
cat('TypeI error(0.05) average_T =', typeI_error2.05, '\n')
cat('TypeI error(0.05) average_Z=PP p_value=', typeI_error3.05, '\n')
cat('TypeI error(0.05) PBT_plugin =', typeI_error4.05, '\n')
cat('TypeI error(0.05) PBT_LRT =', typeI_error5.05, '\n')
```

## Result 2 R Code for Power Simulation Results for $IS_{LIN}$ (F) ($K$=3), Method 2

```
##############################################################################
# This Splus/R Function to compare the power of P values of my ph.D research
# ni = sample size for random variable xi
# Normality
# Ho: P(l'x>0)<=Pi
######################################################################################
general.p_value<-function(n1,n2,n3,n_alter){

 #mu3<-sqrt(sigma3^2+(sigma1^2+sigma2^2)/4)*n_alter+(mu1+mu2)/2
 sigma3<-sqrt(((mu3-(mu1+mu2)/2)/n_alter)^2-(sigma1^2+sigma2^2)/4)

 x1<-rnorm(n1,mu1,sigma1)     # generate n Normal (0,1) random variables
 x2<-rnorm(n2,mu2,sigma2)     # generate n Normal (0,sqrt(10)) random variables
 x3<-rnorm(n3,mu3,sigma3)     # generate n Normal (3,sigma3) random variables

 var1_hat<-var(x1)*(n1-1)/rchisq(num,df=n1-1)
 var2_hat<-var(x2)*(n2-1)/rchisq(num,df=n2-1)
 var3_hat<-var(x3)*(n3-1)/rchisq(num,df=n3-1)

##################
# plug in p_value

a1<-l1^2*var(x1)/n1+l2^2*var(x2)/n2+l3^2*var(x3)/n3
T_p_plugin<-(l1*mean(x1)+l2*mean(x2)+l3*mean(x3))/sqrt(a1)
dem1<-(l1^2*var(x1)/n1)^2/(n1-1)+(l2^2*var(x2)/n2)^2/(n2-1)+(l3^2*var(x3)/n3)^2/(n3-1)
df1<-a1^2/dem1
delta1<-sqrt(l1^2*var(x1)+l2^2*var(x2)+l3^2*var(x3))*qnorm(pi)/sqrt(a1)

p_value.plugin<-1-pt(T_p_plugin, df=df1, ncp=delta1)

##################
# average p_value for T

a2<-l1^2*var1_hat/n1+l2^2*var2_hat/n2+l3^2*var3_hat/n3
b2<-((n1-1)*var(x1)/var1_hat+(n2-1)*var(x2)/var2_hat+(n3-1)*var(x3)/var3_hat)/(n1+n2+n3-3)

T_p_hat<-((l1*mean(x1)+l2*mean(x2)+l3*mean(x3))/sqrt(a2))/sqrt(b2)

delta2<-sqrt(l1^2*var1_hat+l2^2*var2_hat+l3^2*var3_hat)*qnorm(pi)/sqrt(a2)
p_value.T<-1-pt(T_p_hat, df=n1+n2+n3-3, ncp=delta2)
p_value.average_T<-mean(p_value.T)

##################
# average p_value for Z

Z_p_hat<-(l1*mean(x1)+l2*mean(x2)+l3*mean(x3)-
qnorm(pi)*sqrt(l1^2*var1_hat+l2^2*var2_hat+l3^2*var3_hat))/sqrt(a2)

p_value.Z<-1-pnorm(Z_p_hat)
p_value.average_Z<-mean(p_value.Z)
```

```
################################################
# Parametric Bootstrap Test for T_plugin and LRT

T_p1<-Tstar_plug(n1,n2,n3,x1,x2,x3)
T_p2<-Tstar_LRT(n1,n2,n3,x1,x2,x3)

m<-99

mle2<-mle.iteration(n1,n2,n3,x1,x2,x3,qnorm(pi),1e-5)

if (is.matrix(mle2)==FALSE) {
   p_PBT_plug<-NA
   p_PBT_LRT<-NA

} else {
MLE.mu<-mle2[,1]
MLE.sigma<-sqrt(mle2[,2])

x1.boot1<-replicate(m, rnorm(n1, MLE.mu[1], MLE.sigma[1]))
x2.boot1<-replicate(m, rnorm(n2, MLE.mu[2], MLE.sigma[2]))
x3.boot1<-replicate(m, rnorm(n3, MLE.mu[3], MLE.sigma[3]))

# compute test statistics for each resample

T_boot.1<-rep(0,m)
T_boot.2<-rep(0,m)
for (j in 1:m) {
  T_boot.1[j]<-Tstar_plug(n1,n2,n3,x1.boot1[,j],x2.boot1[,j],x3.boot1[,j])
  T_boot.2[j]<-Tstar_LRT(n1,n2,n3,x1.boot1[,j],x2.boot1[,j],x3.boot1[,j])
}

reject1<-ifelse(T_boot.1>=T_p1,1,0)
reject.2<-ifelse(T_boot.2>=T_p2,1,0)
reject2<-reject.2[!is.na(reject.2)]

p_PBT_plug<-(sum(reject1)+1)/(m+1)
p_PBT_LRT<-(sum(reject2)+1)/(length(reject2)+1)
}

#################
# compute the p_value

p_value<-c(p_value.plugin, p_value.average_T,p_value.average_Z, p_PBT_plug, p_PBT_LRT)
p_value
}

##############################################################################
power<-function(n_alter){

B<-200  # get 200 data sets to do simulation

p.dist<-array(rep(0,5*B),c(5,B))
reject.05<-array(rep(0,5*B),c(5,B))
for (i in 1:B){
 p.dist[,i]<-general.p_value(n1,n2,n3,n_alter)
```

```
 reject.05[,i]<-ifelse(p.dist[,i]<=0.05,1,0)
}

re1<-reject.05[1,]
re2<-reject.05[2,]
re3<-reject.05[3,]
re.4<-reject.05[4,]
re4<-re.4[!is.na(re.4)]
re.5<-reject.05[5,]
re5<-re.5[!is.na(re.5)]

typeI_error1.05<-sum(re1)/length(re1)
typeI_error2.05<-sum(re2)/length(re2)
typeI_error3.05<-sum(re3)/length(re3)
typeI_error4.05<-sum(re4)/length(re4)
typeI_error5.05<-sum(re5)/length(re5)

typeI_error.05<-c(typeI_error1.05, typeI_error2.05, typeI_error3.05, typeI_error4.05, typeI_error5.05)
typeI_error.05
}

###########################################################################################
begin.time=Sys.time()
set.seed(6543267)
num<-1000  # get 1000 number of independant var_hat from the chi-square distribution

n1=100
n2=150
n3=120

l1=-0.5
l2=-0.5
l3=1

##########################################################
#(a)equal variance

#mu1<-0
#mu2<-0.5
#sigma1<-1
#sigma2<-1
#sigma3<-1
#mu3<-sqrt(sigma3^2+(sigma1^2+sigma2^2)/4)*qnorm(pi_true)+(mu1+mu2)/2

##########################################################
#(b)sigmasq_i=mu_i+1
mu1<-0
sigma1<-1
mu2<-0.5
sigma2<-sqrt(mu2+1)
mu3<-2
#sigma3<-sqrt(((mu3-(mu1+mu2)/2)/qnorm(pi_true))^2-(sigma1^2+sigma2^2)/4)

###################################################################
pi<-0.75
pi_j<-seq(0.50,pi,by=0.05)
```

```
npai_j<-qnorm(pi_j)

pi_true<-0.75

#m<-c(0.86,0.88,0.90,0.92,0.94,0.96,0.98) #pi=0.85
#m<-c(0.81,0.83,0.85,0.87,0.89,0.91,0.93,0.95,0.98) #pi=0.80
m<-c(0.76,0.79,0.82,0.85,0.88,0.91,0.95,0.98) #pi=0.75
#m<-c(0.66,0.69,0.72,0.76,0.8,0.84,0.88,0.92,0.95,0.98) #pi=0.65
#m<-c(0.56,0.60,0.65,0.70,0.75,0.80,0.85,0.90,0.94,0.98) #pi=0.55

n_alter.m<-qnorm(m,mean=0,sd=1)

mm<-length(m)

power_function<-array(rep(0,5*mm),c(5,mm))
for (j in 1:mm){
  power_function[,j]<-power(n_alter.m[j])
}
end.time=Sys.time()
tt=end.time-begin.time
tt


plot(m,power_function[1,],type="l",lwd=3,ylab="the power",col="blue",xlab="pi")
lines(m,power_function[2,],type="l",lwd=3,col = "red")
lines(m,power_function[3,],type="l",lwd=3,col = "green",lty = "dashed")
lines(m,power_function[4,],type="l",lwd=3,col = "black",lty = "dashed")
lines(m,power_function[5,],type="l",lwd=3,col = "yellow",lty = "dashed")
```