INFERENCE FOR THE *K*-SAMPLE PROBLEM BASED ON PRECEDENCE PROBABILITIES

by

RAJARSHI DEY

B.S., Presidency College Kolkata (University of Calcutta), 2001
M.S., Indian Institute of Technology Kanpur, 2003
M.S., University of Florida, 2005

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Statistics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2011

ABSTRACT

Rank based inference using independent random samples to compare $K \geq 2$ continuous distributions, called the *K-sample* problem, based on precedence probabilities is developed and explored. There are many parametric and nonparametric approaches, most dealing with hypothesis testing, to this important, classical problem. Most existing tests are designed to detect differences among the location parameters of different distributions. Best known and most widely used of these is the F - test, which assumes normality. A comparable nonparametric test was developed by Kruskal and Wallis (1952). When dealing with location-scale families of distributions, both of these tests can perform poorly if the differences among the distributions are among their scale parameters and not in their location parameters. Overall, existing tests are not effective in detecting changes in both location and scale. In this dissertation, I propose a new class of rank-based, asymptotically distribution-free tests that are effective in detecting changes in both location and scale based on precedence probabilities. Let X_i be a random variable with distribution function F_i; $i = 1, 2, ..., K.$ Also, let π be the set of all permutations of the numbers $(1, 2, ..., K)$. Then $P(X_{i_1} < X_{i_2} < ... < X_{i_K})$ is a precedence probability if $(i_1, i_2, ..., i_K) \in \pi$. Properties of these of tests are developed using the theory of *U*-statistics (Hoeffding, 1948).

Some of these new tests are related to volumes under *ROC* (Receiver Operating Characteristic) surfaces, which are of particular interest in clinical trials whose goal is to use a score to separate subjects into diagnostic groups. Motivated by this goal, I propose three new index measures of the separation or similarity among two or more distributions. These indices may be used as “effect sizes”.

In a related problem, Properties of precedence probabilities are obtained and a bootstrap algorithm is used to estimate an interval for them.

INFERENCE FOR THE *K*-SAMPLE PROBLEM BASED ON PRECEDENCE PROBABILITIES

by

RAJARSHI DEY

B.S., Presidency College Kolkata (University of Calcutta), 2001
M.S., Indian Institute of Technology Kanpur, 2003
M.S., University of Florida, 2005

A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Statistics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2011

Approved by:

Major Professor

Paul I. Nelson

ABSTRACT

Rank based inference using independent random samples to compare $K \geq 2$ continuous distributions, called the *K-sample* problem, based on precedence probabilities is developed and explored. There are many parametric and nonparametric approaches, most dealing with hypothesis testing, to this important, classical problem. Most existing tests are designed to detect differences among the location parameters of different distributions. Best known and most widely used of these is the F - test, which assumes normality. A comparable nonparametric test was developed by Kruskal and Wallis (1952). When dealing with location-scale families of distributions, both of these tests can perform poorly if the differences among the distributions are among their scale parameters and not in their location parameters. Overall, existing tests are not effective in detecting changes in both location and scale. In this dissertation, I propose a new class of rank-based, asymptotically distribution-free tests that are effective in detecting changes in both location and scale based on precedence probabilities. Let X_i be a random variable with distribution function F_i; $i = 1, 2, ..., K$. Also, let π be the set of all permutations of the numbers $(1, 2, ..., K)$. Then $P(X_{i_1} < X_{i_2} < ... < X_{i_K})$ is a precedence probability if $(i_1, i_2, ..., i_K) \in \pi$. Properties of these of tests are developed using the theory of *U*-statistics (Hoeffding, 1948).

Some of these new tests are related to volumes under *ROC* (Receiver Operating Characteristic) surfaces, which are of particular interest in clinical trials whose goal is to use a score to separate subjects into diagnostic groups. Motivated by this goal, I propose three new index measures of the separation or similarity among two or more distributions. These indices may be used as “effect sizes”.

In a related problem, Properties of precedence probabilities are obtained and a bootstrap algorithm is used to estimate an interval for them.

Table of Contents

List of Figures

List of Tables

Acknowledgements

Faculty members, family members and friends; if this was just a brief summary, my deepest gratitude and thankfulness would probably have stopped at just that. However, the long journey that was needed to complete the dissertation demands more.

I would first like to thank my strong and always supportive advisor Dr. Paul I. Nelson; without whom this would not have been possible. To be stated properly, I really do not know what would have happened to me if not rescued and persuaded by Dr. Nelson to continue. At times, he had more faith in my ability than anyone else (including me) and this dissertation is the result of his persistence, encouragement and so many painstaking reviews of my novice write-ups.

I would like to thank my committee members Dr. Jim Neill, Dr. Weixing Song, Dr. Robert Burckel and outside chair Dr. Christy Craft for their time, commitment and valuable suggestions.

I sincerely thank the Department of Statistics at the Kansas State University for providing me with many opportunities to fulfill my dream. I am grateful not only to the wonderful teachers and scholars (including but not restricted to Dr. Dallas Johnson, Dr. James Higgins, Dr. Shie-Shien Yang, Dr. Tom Loughin, Dr. Haiyan Wang, Dr. Suzanne Dubnicka, Dr. Leigh Murray, Dr. Abigail Jager and Dr. John Boyer) but also to the staff (particularly, Pam Schierer and Angela Ladner), officemates (Shannon Dick, Matt Hoblin, Will Chernoff, Chris Juarez, Seth Demel, Shu Chen) and friends (Champa Magalla, Nisantha Samarakoon, Shyamalee Kumary, Roshan Serasinghe, Rebhi Bsharat, Robert Poulson, Indu Seetharaman, Nicole Dick and many others) who all have created such a wonderful environment for studying and helped me immensely during my research years.

I would like to express my gratitude for Dr. Bishwanath Das, Dr. Atindra Mohan Gun, Dr. Divakar Sharma, Dr. P C Joshi and Dr. Ronald Randles; who all, in different stages of my education have mentored me to become a statistician. I would also like to thank Dr. Saheli Datta, Sameek Roychoudhury, Subhankar Chowdhury, Soumita Lahiri, Arnab Bhattacharya, Abhrajit Hazra, Dr. Suman Kumar Sen, Dr. Paramita Saha, Dr. Mayukh Samanta, Dr. Sourish Chandra Saha, Dr. Vivekananda Roy, Dr. Sounak

Chakraborty, Dr. Ananya Roy and many others with whom I have enjoyed learning the subject at different points of time.

I must also mention few extraordinary people who were my home away from home in India. I thank Dr. Amitabha Chakrabarti and Dr. Leena Chakrabarti for transporting us back to India whenever needed in the matter of a weekend evening. I thank Dr. Bimal Kumar Paul, Sanghamitra Das, Sanjeev Uruppattur, Swasati Mukherjee and late Dr. Prakash Krishnaswami for their help throughout our stay in Manhattan. I would hereby like to thank Dr. Sohini Dutt, Dr. Debabani Ganguly, Dr. Sankar De, Dr. Sanmitra Barman, Dr. Pankaj Kanti Mandal, Tripti Nanda, Abhishek Rakshit, Achira Mukhopadhyay, Dr. Dipanwita Ray, Santanu Roy, Anuradha Ghosh, Brett Crow and few others who have made the memories in Manhattan more enjoyable.

I express my gratitude to many other friends with whom I have shared a lot of my time starting from my childhood days. I am afraid to mention names which might turn this already long acknowledgement into a really long essay; but contributions of Indraneel Mukherjee, Aban Saha, Dr. Swarnendu Roy, Dr. Sayandev Chatterjee, Dr. Debamalya Banerjee and Sourav Saha must not be forgotten.

No note would be enough to describe how thankful I am to my immediate family. My father Mr. Rabindra Nath Dey, mother Kaberi Dey, only brother Saptarshi Dey and other close family members proudly supported me to make me the first doctorate in my immediate family.

Last, but not the least, I would like to thank my wife, Ayota Banerjee who with her enthusiasm and love had to endure all my tantrums, despair and failures (not to mention statistical theories in the dead of night; although that actually might have helped her sleep) but still was always there for me. Without her sacrifice and indulgence, this dissertation would not have been possible.

I would like to end with expressing my excitement about this dissertation in the words of famous poet Rabindranath Tagore

"*Aakashbhora surjo tara, bishwabhora pran*

Tahari maajhkhaane ami peyechhi mor sthaan

Bishmoye taai jaage amar gaan."

Loosely translated, this means "I am in awe and am grateful that I have found a place in this world. Out of this wonder, my song is born." I sincerely hope so.

Chapter 1: A Motivating Example

Several types of stochastic relationships have been used to test for *a priori* specified ordered alternatives. The example of ordered alternatives that motivates this research comes from the field of medical science.

In particular, consider checking the effectiveness of a diagnosis of a particular disease based on measurements on one or more scales. Often, it is assumed there are only two groups of people: healthy and diseased. However, in some situations, there is also a transition period. If detected in this period, a disease can, in some cases, be cured totally or at least monitored for a longer period of time. One such example is Alzheimer's disease (*AD*), which leads to the death of irreplaceable brain cells. At the beginning stage of this disease, almost no clinical symptoms can be detected, even using the most advanced tests available. But, when someone is close to the threshold of dementia, the person's behavior can suggest that a diagnosis of *AD* lies ahead. It is important that *AD* be detected at this early stage. Although there is now no effective treatment for *AD*, it is hoped that early detection would give a person her/his best chance of avoiding the terrible consequences of this disease when new drugs become available.

Therefore, researchers are looking for a group of tests which can detect *AD* in its early form. To accomplish this goal, requires three diagnostic groups: the healthy people who do not have any symptom of *AD*, the transition group with

subtle hints of *AD* and the group with the disease. Of course, a good diagnostic test should have a discriminating ability among all the groups. A test that has a good discriminating power for only two of the groups may not be very useful.

Generally, most diagnostic tests are based on an observed variable which lies on a continuous or a graded scale. Specifically, consider a score *T* and *ordered* cut points $\{t_i,\ i = 0,1,...,K; t_0 = -\infty, t_K = \infty\}$, where a subject is assigned to diagnostic category k if $t_{k-1} < T \le t_k$, k = 1, 2,…, *K*. Usually *K* is taken to be 2, denoting 'healthy' and 'diseased', and diagnostic accuracy of score *T*, which depends on the cut points used, has historically been based on the probability of an incorrect decision in the healthy population and the probability of a correct decision in the diseased population. It should be noted that the probability of a correct decision in the healthy population is known as *specificity* and the probability of a correct decision in the diseased population is known as *sensitivity*. In this setting, for *K* =2, a perfect test score *T*, if it exists, should have a threshold or cut-off which perfectly discriminates among the diseased and the healthy populations. Typically, T is calibrated so that the first category consists of healthy individuals and increasing values correspond to worsening diagnostic categories. In what follows, the phrase *diagnostic test* refers to a score T and a specific set of cut points.

To study the effectiveness of a diagnostic test based on score *T*, one of the most widely used tools when *K* = 2 is known as the *Receiver Operating Characteristic* (ROC) curve. This curve plots the probability of correctly classifying a diseased

individual in its Y axis and the probability of wrongly classifying a healthy patient in its X axis for each possible choice of the cut point t_1. If an exact diagnostic test based on *T* exists, the *ROC* curve for that test should then go through (0, 1). In reality, that almost never happens. So, a very good test should be as close as possible to (0, 1) in the plot. A very popular index used to measure the discriminating ability of tests is the area or the partial area under a ROC curve. Detailed discussions of ROC curves can be found in Kang (2002) and Hsieh and Turnbull (1996).

In the case of three diagnostic categories, K=3, Mossman (1999) proposed constructing a polyhedral ROC surface and using the *volume under the surface* (VUS) an index to measure the discriminating ability of a diagnostic test. Xiong *et al.* (2006) proposed a similar ROC surface and used the entire volume under the ROC surface as a summary measure to describe the accuracy of a diagnostic test.

The ROC surface proposed by Xiong (2006) assumes normality and represents the probability of correctly recognizing the healthy population, category 1, on the X axis, the probability of correctly recognizing the diseased population, category 2, on the Y axis and the probability of correctly recognizing the intermediate population, category 3, on the Z axis.

Let X_1, X_2 and X_3 denote independent values of a continuous score *T* obtained by subjects from healthy, intermediate and diseased groups, respectively. We will also assume these random variables are independent observations from

$K = 3$ absolutely continuous distribution functions F_1, F_2 and F_3. Thus, if a subject belongs to category j, $T = X_j, j = 1,2,3$ and, as described above for fixed cut points, if score T is: (i) below t_1, assign the subject in the healthy group; (ii) more than t_2, assign the subject in the diseased group and (iii) otherwise assign the subject to the intermediate group. Then, let x be the probability that a randomly selected subject from the healthy group has a test score below t_1, y be the probability that a randomly selected subject from the diseased group has a test score above t_2 and z be the probability that a randomly selected subject from the intermediate group has a test score between t_1 and t_2.

Then, we have the following results.

$$x = F_1(t_1),$$

$$y = 1 - F_3(t_2),$$

$$z = F_2(t_2) - F_2(t_1) = F_2(F_3^{-1}(1-y)) - F_2(F_1^{-1}(x)). \qquad (1.1)$$

We should also note that in this version of an ROC surface, an ideal diagnostic test's ROC surface should pass through (1, 1, 1), which would imply perfect discrimination ability. We should therefore be looking for tests which produce an ROC surface very close to this ideal point.

As shown by Xiong, *et al.* (2006), the VUS for this ROC surface can be expressed as

$$V = \iint_D (F_2(F_3^{-1}(1-y)) - F_2(F_1^{-1}(x)))dxdy$$

$$= P(X_1 < X_2 < X_3) \quad . \tag{1.2}$$

Integration in (1.2) is over the domain $D = \{(x,y) \mid 0 \le x \le 1, 0 \le y \le 1 - F_3(F_1^{-1}(x))\}$, since $t_1 < t_2$.

To further assess the effectiveness of any diagnostic test, we proceed by constructing appropriate hypotheses about the distributions of X_1, X_2, X_3 . If a diagnostic test does not have any discrimination power, X_1, X_2, X_3 are identically distributed, all 6 orderings of X_1, X_2, X_3 are then equally likely and $V = 1/6$. This observation motivates us to explore the world of precedence orderings which deals with the distributions of order statistics when marginal distributions are not identical. Probabilities of these orderings are called *precedence probabilities*. The observation that the VUS can be represented as a probability also suggests additional problems such as testing hypotheses for ordered alternatives and constructing confidence intervals for V . We will later use data consisting of fourteen test scores on each of three known categories of *AD* subjects to illustrate the methods developed in this study.

Consider a generalization of this problem. Suppose that, we are interested in a medical test having outcome T that classifies patients into one of the K groups. Let the values of T be such that group K tends to score higher than

group $(K-1)$ and so on with group 1 scoring the lowest. In that case, we can think about a *ROC* manifold with a point on this manifold representing the probability of getting selected correctly into a group. Instead of looking at VUS, we will then be looking at a *Hypervolume Under Manifold* (HUM), which equals $P(X_1 < X_2 < ... < X_K)$.

In sum, motivated by this example, part of my research concerns constructing nonparametric tests of the hypothesis of randomness

$$H_0 : F_1 = F_2 = ... = F_K \ , \tag{1.3}$$

versus ordered alternatives motivated by (1.2) and discussed in full in Chapters 3, 8 and 11. The tests I develop are based on U-Statistics, a class of statistics described in the next chapter. The performance of this new test is described in Chapter 11.

Moving beyond this motivating example, in Chapters 8 and 9 I provide new rank-based tests for the classical K-sample problem, where we test the equality of K distributions based on independent random samples. These new tests are shown to provide significant improvements to classical tests. In particular, these new tests have the ability to detect changes in location, in scale or both, whereas most existing tests are powerful in detecting changes in only one of these features. Since the K-sample problem is one of the oldest and most important problems in the statistical universe, the proposed tests are a valuable addition to the existing literature.

As a practical matter, no two distributions are identical. So, instead of testing for equality, researchers are often more interested in assessing just how different distributions are. Therefore, constructing measures of distance among K distributions is an important problem, especially for $K \geq 3$. Here, I propose new measures which can be used to assess what is commonly called *effect size*.

In sum, in this thesis, I explore and develop properties of precedence probabilities and use these results to create inference procedures for the K-sample problem that perform well in detecting important alternatives that are not often considered and will provide researchers with important new tools.

Chapter 2: U-Statistics

Hoeffding (1948) developed the basic theory of U-statistics. Detailed discussions about this topic are found in Lee (1990), Serfling (2001) and Randles and Wolfe (1979).

Let $\mathbf{P} = \{P\}$ be a family of probability measures on an arbitrary measurable space and let θ be a real-valued function defined on $\mathbf{P}$. The parameter $\theta(P) = \theta$ is said to be *estimable of degree* r for $\mathbf{P}$ if r is the smallest sample size for which there exists a real-valued measurable function $h(x_1, x_2, ..., x_r)$ such that

$$E_P(h(X_1, X_2, ..., X_r)) = \theta \qquad \text{for all } P \in \mathbf{P}, \tag{2.1}$$

where $\{X_i\}$ are independent, identically distributed according to P.

The function $h(.)$ defined in (2.1) is known as the *kernel* of the associated parameter θ.

The *kernel*, without loss of generality, may be assumed to be a symmetric function of its arguments. This is because if h^* is an unbiased estimator of θ, then so is the symmetric function

$$h(x_1, x_2, ..., x_r) = \frac{1}{r!} \sum_{\pi \in \Pi_r} h^*(x_{\pi_1}, x_{\pi_2}, ..., x_{\pi_r}) \tag{2.2}$$

where the summation is over the group $\prod_r$ of all permutations of the first r integers. So, henceforth the symmetry assumption on h will be in effect.

For a symmetric *kernel* $h(.)$ and a sample $X_1, X_2, ..., X_n$ of independent, identically distributed random variables each of whose distributions is determined by P, with $n > r$, a *U*-statistic is defined as

$$U_n = \frac{1}{C_r^n} \sum_{\beta \in B} h(X_{\beta_1}, X_{\beta_2}, ..., X_{\beta_r}) , \tag{2.3}$$

where $B = \{\beta \mid \beta$ is one of the C_r^n unordered subsets of r positive integers chosen without replacement from the set $\{1, 2, ..., n\}$.

The sample mean is a *U*-statistic of degree 1 with $h(x) = x$ for estimating the population mean. In particular, the sample mean is the U-statistic estimator of the population mean. Likewise, the sample variance is the *U*-statistic estimator of the population variance. The statistic used in the sign test is also a U-statistic. The Wilcoxon signed-rank test statistic can also be written as a combination of two different *U*-statistics.

A very desirable property of a *U*-statistic is that of being a minimum-variance unbiased estimator. Specifically, if **P** includes *all* continuous distributions, a *U*-statistic can be shown to be the unique minimum-variance unbiased estimator (UMVUE) of the associated parameter θ.

As presented in Randles and Wolfe (1979), the variance of the *U*-statistic in (2.3) is given by

$$Var[U_n] = \frac{1}{C_r^n} \sum_{l=1}^{r} C_l^r C_{r-l}^{n-r} \tau_l \quad , \tag{2.4}$$

where

$$\tau_l = Cov\{h(X_{\beta_1}, X_{\beta_2}, ..., X_{\beta_r}), h(X_{\alpha_1}, X_{\alpha_2}, ..., X_{\alpha_r})\}, \tag{2.5}$$

and $\beta = \{\beta_1, \beta_2, ..., \beta_r\}$ and $\alpha = \{\alpha_1, \alpha_2, ..., \alpha_r\}$ are subsets of the first n positive integers having exactly l integers in common.

If $\tau_r < \infty$ and $\tau_1 > 0$, $\sqrt{n}(U_n - \theta)$ has a limiting normal distribution with mean 0 and variance $r^2\tau_1$, as stated and shown in Randles and Wolfe (1979).

We can extend this concept to the more general case of K distributions. Let $X_{ij}; j = 1, 2, ..., n_i; i = 1, 2, ..., K$ be independent observations with all the observations in $X_i = \{X_{ij}; j = 1, 2, ..., n_i\}$ being generated from distribution function (d.f.) F_i.

A parameter θ is said here to be *estimable of degree* $(r_1, r_2, ..., r_K)$ for distributions $(F_1, F_2, ..., F_K)$ in some family of probability measures $\mathbf{P}$, if $(r_1, r_2, ..., r_K)$ are the smallest possible sample sizes for which an unbiased estimator for θ exists for every $(F_1, F_2, ..., F_K) \in \mathbf{P}$. A $K-$sample symmetric *kernel* for θ can be obtained by generalizing the 1-sample approach mentioned in (2.3). Specifically, let $h(x_{11}, x_{12,}..., x_{1r_1}; x_{21}, x_{22,}..., x_{2r_2}; ...; x_{K1}, x_{K2,}..., x_{Kr_K})$ be a function that is symmetric in the arguments $(x_{i1}, x_{i2,}..., x_{ir_i})$ for each $i = 1, 2, ..., K$. Then, $h(.)$ is a $K-$sample symmetric kernel for θ if for all $(F_1, F_2, ..., F_K) \in \mathbf{P}$,

$$E_{(F_1,F_2,...,F_K)}(h(X_{11},X_{12,}...,X_{1r_1};X_{21},X_{22,}...,X_{2r_2};...;X_{K1},X_{K2,}...,X_{Kr_K}))=\theta(F_1,F_2,...,F_K)\,. \quad (2.6)$$

For an estimable parameter θ of degree $(r_1,r_2,...,r_K)$ and a symmetric kernel $h(.)$ with $n_i \geq r_i; i=1,2,...,K$; a *U*-statistic is defined as

$$U_{n_1n_2...n_K}=\frac{1}{C_{r_1}^{n_1}C_{r_2}^{n_2}...C_{r_K}^{n_K}}\sum_{\beta_1\in B_1}\sum_{\beta_2\in B_2}...\sum_{\beta_K\in B_K}h(X_{1\beta_{11}},X_{1\beta_{12},}...,X_{1\beta_{1r_1}};X_{2\beta_{21}},X_{2\beta_{22},}...,X_{2\beta_{2r_2}};...;X_{K\beta_{K1}},X_{K\beta_{K2},}...,X_{K\beta_{Kr_K}})$$

(2.7)

where $\beta_i=(\beta_{i1},\beta_{i2},...,\beta_{ir_i})$ and B_i is the collection of all subsets of r_i positive integers chosen without replacement from $(1,2,...,n_i)$ for $i=1,2,...,K$.

Asymptotic normality can be obtained for the *U*-statistic defined in (2.7). Define for $i=1,2,...,K$,

$$H_{i\beta}=h(X_{1\beta_{11}},X_{1\beta_{12},}...,X_{1\beta_{1r_1}};X_{2\beta_{21}},X_{2\beta_{22},}...,X_{2\beta_{2r_2}};...;X_{K\beta_{K1}},X_{K\beta_{K2,}}...,X_{K\beta_{Kr_K}}),$$
$$H_{i\gamma}=h(X_{1\gamma_{11}},X_{1\gamma_{12},}...,X_{1\gamma_{1r_1}};X_{2\gamma_{21}},X_{2\gamma_{22},}...,X_{2\gamma_{2r_2}};...;X_{K\gamma_{K1}},X_{K\gamma_{K2,}}...,X_{K\gamma_{Kr_K}})\,,$$

(2.8)

where the two sets $(\beta_{j1},\beta_{j2},...,\beta_{jr_j})$ and $(\gamma_{j1},\gamma_{j2},...,\gamma_{jr_j})$ have no common integers when $j\neq i$ and exactly one common integer when $j=i$.

Let us also define $\rho_i=Cov(H_{i\beta},H_{i\gamma})$ for $i=1,2,...,K$.

Then, Randles and Wolfe (1979) provide us with the following result.

Lemma 2.1: Let $U_{n_1 n_2 \ldots n_K}$ be a K sample U-statistic for parameter θ of degree $(r_1, r_2, \ldots, r_K)$. If $N = \sum_{i=1}^{K} n_i$ and $\lim_{N\to\infty}(n_i / N) = \lambda_i, 0 < \lambda_i < 1; \text{for} \quad i = 1, 2, \ldots, K$ and if $E(h^2(X_{11}, X_{12,}\ldots, X_{1r_1}; X_{21}, X_{22,}\ldots, X_{2r_2}; \ldots; X_{K1}, X_{K2,}\ldots, X_{Kr_K})) < \infty$, then $\sqrt{N}(U_{n_1 n_2 \ldots n_K} - \theta)$ has a limiting normal distribution with mean 0 and variance

$$\sigma^2 = \sum_{i=1}^{K} \frac{r_i^2}{\lambda_i}\rho_i, \tag{2.9}$$

provided $\sigma^2 > 0$.

Since my research deals with a K sample case, this result provides a guideline for later derivations. In the next chapter, I will describe different types of ordering among distributions.

Chapter 3: Ordered Alternatives

Some of my research concerns alternative hypotheses that involve some *a priori* ordering among $K \geq 2$ distributions. The work of Xiong, *et al.* (2006) provides a practical motivation, as discussed in chapter 1. I will start by defining four well- known orderings of random variables. These are: *likelihood ratio ordering*, *uniform stochastic (hazard rate) ordering*, *stochastic ordering* and *precedence ordering*. Ross (1996) provides a detailed discussion of all these orderings except the last one.

Let F_i denote the absolutely continuous distribution function of random variable X_i and let f_i denote its density, $i = 1, 2, ..., K$. We say X_K is larger than X_{K-1} is larger than X_{K-2} ……is larger than X_1 in the sense of *likelihood ratio ordering*, and write

$$X_K \underset{LR}{\geq} X_{K-1} \underset{LR}{\geq} ... \underset{LR}{\geq} X_2 \underset{LR}{\geq} X_1, \text{ if } \frac{f_j(x)}{f_{j+1}(x)} \geq \frac{f_j(y)}{f_{j+1}(y)} \tag{3.1}$$

for $j = 1, 2, ..., K-1$ and for all $x \leq y$ with strict inequality being true for at least one pair of (x, y) for each pair of the random variables (X_i, X_{i+1}) where the ratios in (3.1) are defined.

The *hazard rate* λ_j of a continuous random variable X_j is defined as

$$\lambda_j(t) = \frac{f_j(t)}{1 - F_j(t)} \tag{3.2}$$

We say X_K is larger than X_{K-1} is larger than X_{K-2}……is larger than X_1 in the sense of *hazard rate ordering*, and write

$$X_K \underset{HR}{\geq} X_{K-1} \underset{HR}{\geq} \ldots \underset{HR}{\geq} X_2 \underset{HR}{\geq} X_1, \text{ if } \lambda_1(t) \geq \lambda_2(t) \geq \ldots \geq \lambda_K(t) \tag{3.3}$$

for all t where the hazard rate is defined , with strict inequality for at least one $t = t_i$ for each pair of random variables (X_i, X_{i+1}).

We say X_K is larger than X_{K-1} is larger than X_{K-2}……is larger than X_1 in the sense of *stochastic ordering*, and write

$$X_K \underset{ST}{\geq} X_{K-1} \underset{ST}{\geq} \ldots \underset{ST}{\geq} X_2 \underset{ST}{\geq} X_1, \text{ if } F_1(x) \geq F_2(x) \geq \ldots \geq F_K(x) \tag{3.4}$$

for all x, with strict inequality for at least one $x = x_i$ for each pair of random variables (X_i, X_{i+1}).

Note that for a shift alternative model $F_j(x) = F_1(x - \theta_j)$, $j = 2, 3, \ldots, K$ with $\theta_2 < \theta_3 < \ldots < \theta_K$, the stochastic ordering given in (3.4) prevails. Behnen and Neuhaus (1989) developed an adaptive rank-type test for stochastic ordering designed to detect stochastic ordering based on independent random samples.

If $\{X_i\}$ are jointly independent, we say X_K is larger than X_{K-1} is larger than X_{K-2} ……is larger than X_1 in the sense of *precedence ordering* and write

$$F_1 \prec F_2 \prec \ldots \prec F_K \quad \text{if } \Pr(X_1 < X_2 < \ldots < X_K) \geq \frac{1}{K!}. \tag{3.5}$$

It can be shown that stochastic ordering is the weakest relationship among the first three relationships described. In fact, *likelihood ratio ordering* implies *hazard rate ordering* and in turn, *hazard rate ordering* implies stochastic ordering.

As argued by Arcones, *et al.* (2002), even stochastic ordering is too strong an assumption for practical use since it assumes uniform domination of one distribution function over another. I provide the following example to describe this.

Example 3.1: Consider two independent random variables X_1 and X_2 where X_1 is observed from a uniform $(0,1)$ distribution while X_2 has a distribution function F_2 defined as the following:

$$F_2(x) = 3x^2 \quad , 0 \leq x \leq \frac{1}{3}$$

$$= 3x - 1.5x^2 - 0.5 \quad , \ \frac{1}{3} < x \leq 1. \tag{3.6}$$

It is easy to show that, $\Pr(X_2 > X_1) = \frac{2}{3}$.

But, since $F_2(x) > F_1(x)$ for $\frac{1}{3} < x \leq 1$, X_2 is not stochastically larger than X_1.

Arcones, *et al.* (2002) also provides some examples of the use of stochastic precedence in the literature of statistics. Among other notable uses of stochastic precedence is to develop an approximate solution to the classical Behrens-Fisher problem.

I will focus on constructing new nonparametric tests for the *K*-sample problem based on stochastic precedence. First, I show that precedence ordering is weaker than stochastic ordering. The following theorem relates stochastic and precedence ordering.

Theorem 3.1: Let $X_K \underset{ST}{\geq} X_{K-1} \underset{ST}{\geq} \ldots \underset{ST}{\geq} X_2 \underset{ST}{\geq} X_1$ be stochastically ordered, independent random variables with X_i having continuous distribution function F_i and p.d.f. f_i, $i = 1, 2, \ldots, K$.

Then,

$$\Pr(X_1 < X_2 < \ldots < X_K) > \frac{1}{K!} \quad . \tag{3.7}$$

Proof: The assumption of stochastic ordering implies that for at least one $j_0 \in \{1, 2, ..., K-1\}$, $\exists$ some $x_0 \in (-\infty, \infty)$ such that $F_{j+1}(x_0) < F_j(x_0)$.

Without loss of generality, let us assume $f_j(x_0) > 0$. Continuity of the distribution functions then yields that

$$F_{j+1}(x_0) = \underset{\delta \to 0}{Lim}\ F_{j+1}(x_0 + \delta) < \underset{\delta \to 0}{Lim}\ F_j(x_0 + \delta) = F_j(x_0). \tag{3.8}$$

Let $F_j(x_0) - F_{j+1}(x_0) > 2\varepsilon > 0$.

Then, for all sufficiently small $\delta > 0$, for all $x \in (x_0 - \delta, x_0 + \delta)$,

$$F_j(x) - F_{j+1}(x) > \varepsilon > 0. \tag{3.9}$$

Recall that for a continuous distribution function $F(x)$ and density function $f(x)$,

$$\int_a^b F^r(x) f(x) dx = \int_a^b F^r(x) dF(x) = \frac{F^{r+1}(b) - F^{r+1}(a)}{(r+1)} \text{ for } r = 0, 1, 2, ... \tag{3.10}$$

First consider the case $K = 2$. If (3.4) holds,

$$\Pr(X_1 < X_2) = \iint_{x_1 < x_2} f_1(x_1) f_2(x_2) dx_1 dx_2$$

$$= \int_{-\infty}^{\infty} \left(\int_{x_1}^{\infty} f_2(x_2) dx_2 \right) f_1(x_1) dx_1$$

$$= \int_{-\infty}^{\infty} \left(\int_{F_2(x_1)}^{1} dF_2(x_2) \right) f_1(x_1)dx_1$$

$$= \int_{-\infty}^{\infty} \big(1 - F_2(x_1)\big) f_1(x_1)dx_1 \qquad \text{(Using (3.10))}$$

$$= 1 - \int_{-\infty}^{\infty} F_2(x_1) f_1(x_1)dx_1 \qquad (3.11)$$

Also, $\int_{-\infty}^{\infty} F_2(x_1) f_1(x_1)dx_1 = \int_{-\infty}^{x_0-\delta} F_2(x_1) f_1(x_1)dx_1 + \int_{x_0-\delta}^{x_0+\delta} F_2(x_1) f_1(x_1)dx_1 + \int_{x_0+\delta}^{\infty} F_2(x_1) f_1(x_1)dx_1$

$$\leq \int_{-\infty}^{x_0-\delta} F_1(x_1) f_1(x_1)dx_1 + \int_{x_0-\delta}^{x_0+\delta} F_2(x_1) f_1(x_1)dx_1 + \int_{x_0+\delta}^{\infty} F_1(x_1) f_1(x_1)dx_1 \qquad \text{(Using (3.4))}$$

$$< \int_{-\infty}^{x_0-\delta} F_1(x_1) f_1(x_1)dx_1 + \int_{x_0-\delta}^{x_0+\delta} F_1(x_1) f_1(x_1)dx_1 + \int_{x_0+\delta}^{\infty} F_1(x_1) f_1(x_1)dx_1 \qquad \text{(Using (3.9))}$$

$$= \int_{-\infty}^{\infty} F_1(x_1) dF_1(x_1)$$

$$= \frac{1}{2} \qquad \text{(Using (3.10))} \qquad (3.12)$$

Applying (3.12) in (3.11), $\Pr(X_1 < X_2) > \frac{1}{2}$.

The verification of (3.7) for K > 2 follows similarly.

Let us note that when $F_1 = F_2 = ... = F_K$, $\Pr(X_1 < X_2 < ... < X_K) = \frac{1}{K!}$.

Using this theorem as a motivation, I will consider (among other tests of interest) the following hypothesis

$$H_0 : F_1 = F_2 = ... = F_K$$

Vs.

$$H_A : \Pr(X_1 < X_2 < ... < X_K) > \frac{1}{K!}.$$

(3.13)

When $K = 2$, for most well-known one-parameter location families of distributions, we believe that stochastic precedence implies stochastic ordering. Unfortunately, for higher values of K this relationship between stochastic ordering and stochastic precedence is not yet understood.

In the next chapter, I review the relevant literature.

Chapter 4: Literature Review

As mentioned earlier, one of the most widely used and discussed problems in statistics is the comparison of distributions $\{F_i : i = 1,2,...,K\}$ based on independent random samples $\{X_{ij};\ i = 1,2,...,K, j = 1,2,...,n_i\}$. Nonparametric hypothesis testing in this setting most often tests $H_0 : F_1 = F_2 = \cdots = F_K$ vs. the non-directional alternative H_a: *at least two of the distributions differ*. Most of the literature available in this area concerns the two-sample case.

To begin our discussion, consider a set of K, independent random variables $(X_1, X_2, ..., X_K)$ with continuous probability density functions (p.d.f.) $(f_1, f_2, ..., f_K)$ and distribution functions $(F_1, F_2, ..., F_K)$. Also consider the rank vectors **R** = $\{\underline{R}_i;\ i = 1,2,...,K\}$, where the elements of $\underline{R}_i$ are the ranks of $\{X_{ij}, j = 1,2,...,n_i\}$ in the pooled N dimensional data vector **X** = $\{X_{ij}\}$.

An important class of test statistics designed to test equality of all elements in $\{F_i; i = 1,2,...,K\}$ is the class of linear rank test statistics. The general idea for all these tests came from two-sample problems which were later extended to *K*-sample problems. A statistic of the form $S_i = \sum_{j=1}^{N} c_{ij} a(R_j)$ is known as a *simple linear rank statistic* for the sample taken from *i*-th population, where R_j is the

rank of the j-th observation, $a(R_j)$ is a score based on that rank, c_{ij} is an indicator variable denoting whether the j-th observation belongs to the i-th population, and N is the total sample size. The idea is to calculate the score S_i for each group with sample size n_i and then to create a one-way analysis of variance (ANOVA)-like test statistic of the form

$$C = \frac{\left(\sum_{i=1}^{K} (S_i - E_{H_o}(S_i))^2 / n_i \right)}{\left(\sum_{j=1}^{N} (a(R_j) - \bar{a})^2 / (N-1) \right)}, \tag{4.1}$$

where $E_{H_0}(S_i) = E(S_i)$ when H_0 is true and $\bar{a}$ is the average of all $a(R_j)$ s.

Choosing different scores $a(R_j)$ leads to different test statistics. $a(R_j)$ is generally chosen to be real-valued and non-decreasing in its argument. Most well known among these linear rank tests is the *Kruskal-Wallis* (KW) test. The test is due to Kruskal and Wallis (1952). A simple score choice of $a(R_j) = R_j$ in (3.1) leads to this test. The test statistic can be written as

$$Kw = \frac{\left(\sum_{i=1}^{K} (\bar{R}_i - \bar{R}))^2 / n_i \right)}{\left(\sum_{j=1}^{N} (R_j - \bar{R})^2 / (N-1) \right)}. \tag{4.2}$$

In the above equation, $\bar{R}_i$ is the average rank of the sample generated from the i-th population and $\bar{R}$ is the average rank of the pooled sample.

Another popular test is the median test proposed by Brown and Mood (1951). They used the median of the pooled observations to create a test statistic based on a chi-square test applied to a $K \times 2$ contingency table to test the equality of distributions against the two-sided alternative that at least two distributions differ. If $a(R_j) = 1$ for $R_j > \frac{N+1}{2}$ and 0 otherwise, we end up with a test equivalent to the median test which is powerful for distinguishing symmetric distributions with heavy tails.

Choosing $a(R_j) = \Phi^{-1}\left(\frac{R_j}{n+1}\right)$, where Φ is the cumulative distribution function of a standard normal distribution leads to the Van der Waerden test, which is powerful for testing for the equality of normal distributions.

Choosing $a(R_j) = \sum_{i=1}^{R_j} \frac{1}{n-i+1} - 1$ leads to the Savage test (log rank test) which is powerful to detect location shifts in extreme value distributions. However, Behnen and Neuhaus (1989) claim that the last three procedures can have low power when used with data generated from distributions other than the ones they were designed for.

A recent nonparametric contribution to the *K*-sample problem can be found in Rizzo and Szekely (2010), who use distances between observations to create a new test for the equality of distributions. Their test is designed to detect the non-directional alternative $H_a : F_i \neq F_j$ for at least one pair (i, j). The distances are calculated based on a Gini mean-distance statistic and a Euclidean norm.

Their method depends on partitioning the total dispersion of the samples, quite like the way the traditional F-test partitions the total variance in ANOVA and is known as the DISCO test.

There are some nonparametric tests designed to detect specific types of alternatives. For example, let $v^{(i)}$ be the number of *K*-tuples that can be formed by choosing one observation from each sample such that the observation from the *i*-th population is the least. Specifically,

$$v^{(i)} = \sum_{j=1}^{n_i} \prod_{r \neq i} \#\{x_{rs} > x_{ij}, s = 1, 2, ..., n_r\}.$$

Bhapkar (1961) used the total number of such *K*-tuples that can be formed as the basis of a test designed to detect a shift alternative.

In life data analysis, rank tests can be used for comparing distributions when the data is right-censored. Prentice (1978) suggested an outline to define these rank tests. Let us consider L distinct observed log lifetimes $z_{(1)} < z_{(2)} < ... < z_{(L)}$ for the pooled sample. In addition, suppose there are u_{il} censoring times from the i -th population falling into the interval $[z_{(l)}, z_{(l+1)})$, for $l = 0, 1, 2, ..., L$, where for convenience we define $z_{(0)} = 0$ and $z_{(L+1)} = \infty$.

Prentice (1978) proposed the use of a score statistic that has components of the form

$$S_i = \sum_{l=1}^{L} (c_{il} a_l + u_{il} b_l) \tag{4.3}$$

for the sample from i-th population, where c_{il} is an indicator variable denoting whether the l-th observation (ordered) belongs to the i-th population, a_l is the score of the sample whose lifetime ended at $z_{(l)}$ and b_l is the score of individuals whose lifetimes are censored in $[z_{(l)}, z_{(l+1)})$. Further, define d_{il} to be number of deaths at $z_{(l)}$ from the i-th population, m_{il} to be the number of individuals from the i-th population at risk just prior to $z_{(l)}$. Also, define $\sum_{i=1}^{K} m_{il} = m_l$ and $\sum_{i=1}^{K} d_{il} = d_l$. Then, the following choice of scores in (3.3) leads to the log rank test suggested by Peto and Peto (1972) and Prentice (1978).

$$a_l = \sum_{j=1}^{l} \left(\frac{1}{m_j} - 1 \right)$$

$$b_l = \sum_{j=1}^{l} \frac{1}{m_j} \qquad \text{for} \quad l = 1, 2, ..., L.$$

Gehan (1965 a, b) and Breslow (1970) suggested the following scores for a generalized Wilcoxon test with censored data.

$$a_l = \frac{l - m_l}{N+1},$$

$$b_l = \frac{l}{N+1}, \quad \text{for} \quad l = 1, 2, ..., L.$$

For the K-sample problem, however, testing against a directional alternative has not been discussed as much as the non-directional alternative. Whitney

(1951) proposed an extension of his two-sample test to the three-sample problem. He considered an ordered alternative (X_1 is stochastically larger than X_2, which in turn is stochastically larger than X_3) and also an umbrella alternative (X_1 is stochastically larger than both X_2 and X_3). The test statistic is based on the joint distribution of the pair of Mann-Whitney-Wilcoxon U-Statistic for X_1 and X_2 and the U-statistic for X_1 and X_3.

The *Jonckheere-Terpstra* (JT) test (1954) is the most well-known test which tests for a stochastically ordered alternative for the *K*-sample problem and is based on another well-known nonparametric test statistic, namely the Mann-Whitney-Wilcoxon U-Statistic designed for the 2-sample problem. It tests the alternative hypothesis $H_a : F_1(x) < F_2(x) < \ldots < F_K(x)$ for all x. Let, w_{ij} be the number of observations in the previous $(i-1)$ samples that are smaller than X_{ij}. The JT test statistic is defined as,

$$J = 2\sum_{i=1}^{K}\sum_{j=1}^{n_i} w_{ij} - \sum_{i=1}^{K-1}\sum_{j=i+1}^{K} n_i n_j \ . \tag{4.4}$$

More detailed discussions of nonparametric tests for the *K*-sample problem can be found in Hollander and Wolfe (1999), Hajek, Sidak and Sen (1999) and Lawless (2003).

Chapter 5: How Precedence Probability Behaves

The alternative hypothesis mentioned in (3.13) deals with a particular precedence probability, which could be termed as the probability of observing one observation from each underlying distribution in a particular order (not in terms of occurrence, rather in terms of the observation itself).

In this chapter, I study the behavior of precedence probabilities in a variety of settings, starting with normal distributions and then moving on to consider heavy-tailed and non-symmetric distributions. Let $\xi_i = (\zeta_{i1}, \zeta_{i2}, ..., \zeta_{iK}); i = 1, 2, ..., K!$ denote the permutations of the first K integers and $p(\xi_i) = P(X_{\xi_{i1}} < X_{\xi_{i2}} < \cdots < X_{\xi_{iK}})$ the corresponding precedence probabilities. Take, without loss of generality the precedence probability $p(\xi_1) = P(X_1 < X_2 < \cdots < X_K)$ to be the main parameter of interest throughout this chapter. This particular precedence probability is also referred to as the preferred precedence. Another parameter of interest is $p_{\max} = Max\{p(\xi_i); i = 1, 2, ..., K!\}$, which might be interpreted as the probability of the most likely ordering selected by nature. Recall that for K continuous random variables, there are $K!$ different precedence probabilities and we always assume $\{X_i; i = 1, 2, ..., K\}$ to be jointly independent and continuous.

The behavior of precedence probabilities is not a very new issue and has been studied under different names for $K = 2$. However, when the number of

underlying distributions is more than 2, the behavior of precedence probabilities has not been studied in detail. Their behavior when the underlying distributions are identical is, of course, simple to describe. Otherwise, the behavior of precedence probabilities has to be investigated on a case by case basis. Here, I chose to study the way precedence probabilities behave when all the distributions are: normal, a symmetric case; logistic, a symmetric distribution with heavy tails; exponential, an asymmetric distribution with a varying left support point; extreme value, an asymmetric distribution. I begin with new results related to the behavior of precedence probabilities.

Theorem 5.1: For any two continuous, independent random variables X_1 and X_2, $P(X_2 < X_1) = P(X_1 < X_2) = 0.5$ if and only if the median of $(X_1 - X_2)$ is 0.

Proof: Suppose, the median of $(X_1 - X_2)$ is 0.

Then, $P(X_1 - X_2 < 0) = P(X_1 - X_2 > 0)$ = 1- $P(X_1 - X_2 < 0)$ = ½.

Conversely, suppose that $P(X_2 < X_1) = P(X_1 < X_2) = 0.5$.

Then, $P(X_2 - X_1 < 0) = P(X_1 - X_2 < 0)$, which completes the proof.

Theorem 5.2: If for any set of independent, continuous random variables $\{X_i; i = 1,2,3\}$; $P(X_1 < X_2) > 0.5$ and $P(X_2 < X_3) > 0.5$;

then $P(X_1 < X_2 < X_3) > P(X_3 < X_2 < X_1)$.

Proof: Since the hypotheses imply that $P(X_2 < X_1) < 0.5$ and $P(X_3 < X_2) < 0.5$, it follows that $P(X_1 < X_2) > P(X_2 < X_1)$

and hence that

$$P(X_1 < X_2 < X_3) + P(X_1 < X_3 < X_2) + P(X_3 < X_1 < X_2) >$$

$$P(X_2 < X_1 < X_3) + P(X_2 < X_3 < X_1) + P(X_3 < X_2 < X_1). \quad (5.1)$$

In a similar manner,

$$P(X_1 < X_2 < X_3) + P(X_2 < X_1 < X_3) + P(X_2 < X_3 < X_1) >$$

$$P(X_3 < X_1 < X_2) + P(X_1 < X_3 < X_2) + P(X_3 < X_2 < X_1). \quad (5.2)$$

Adding equations (5.1) and (5.2), we obtain the desired result, which completes the proof.

5.1 *The Normal Distribution*

I begin by studying the behavior of $p(\xi_1) = P(X_1 < X_2 < ... < X_K)$ when $X_i \sim N(\mu_i, \sigma_i^2)$ for $i = 1, 2, ..., K$. In this case, it is particularly helpful to transform a precedence ordering to a statement that the successive differences $W_i \equiv X_i - X_{i+1}$, $i = 1, 2, ..., K-1$; are negative. Specifically, we have that

$$p(\xi_1) = P(W_1 \le 0, W_2 \le 0, ..., W_{K-1} \le 0) \tag{5.3}$$

and $\mathrm{W} = (W_1, W_2, ..., W_{K-1})' \sim N_{K-1}(\boldsymbol{\varphi}, \Sigma)$. The i-th element of $\boldsymbol{\varphi}$ is $(\mu_i - \mu_{i+1})$; $i = 1, 2, ..., (K-1)$ and Σ is the $(K-1)\times(K-1)$ matrix with i-th diagonal element equal to $(\sigma_i^2 + \sigma_{i+1}^2)$ and (i, j)-th off diagonal element equal to $-\sigma_i^2$ if $j = i-1$ and $-\sigma_{i+1}^2$ if $j = i+1$. The other elements of Σ are zeros.

Note that for $K = 2$, $p(\xi_1) = 0.5 = 1/K!$ even if the variances differ but the means are equal (from Theorem 5.1). Now, I focus on the case K=3. Studying the effects of differences among normal distributions on $p(\xi_1)$ is pretty intuitive if the three distributions differ only in their means or only in their variances. However, when both the means and variances differ, the behavior of $p(\xi_1)$ can be very unintuitive and complicated.

For this study, I used the software package *R* to calculate exact precedence probabilities under normality. I will consider the cases when the differences are only among the location parameters, only among the scale parameters and when the differences are both in location and scale parameters.

Case 1

We assume $X_i \sim N(\mu + (i-1)d, \sigma^2); i = 1, 2, 3$ and $d > 0$. In this case, it is evident that $p(\xi_1)$ and $p_{\max}$ are equal and that $p(\xi_1)$ increases monotonically as d increases.

It is also noteworthy that in this scenario, $p_{\max}$ asymptotes at its maximum possible value of 1. A detailed analysis is presented in Table 5.1.1 and Plot 5.1.1, where, without loss of generality, σ^2 is taken to be 1.

Case 2

Let $X_1, X_2 \sim N(\mu, \sigma^2)$ and $X_3 \sim N(\mu + d, \sigma^2)$.

Again, $p(\xi_1)$ and $p_{\max}$ are equivalent and overall, behave very similar to Case 1. However, $p_{\max}$ has an upper bound here of 0.5, instead of 1. This can be explained by the fact that $P(X_2 < X_1 < X_3)$ is the same as $p(\xi_1)$ since the event that X_2 is less than X_1 is as likely as X_1 being less than X_2. It should also be noted that both $P(X_2 < X_1 < X_3)$ and $p(\xi_1)$ equal $p_{\max}$. This case is presented in detail in Table 5.1.2 and Plot 5.1.2, with both X_1 and X_2 being standard normal.

Case 3

We assume, $X_i \sim N(\mu, \sigma_i^2); i = 1,2,3$ with $\sigma_{i+1}^2 = \sigma_i^2 + s; i = 1,2; s > 0$. In this scenario, $p(\xi_1)$ can easily be calculated using a result on orthant probabilities of

trivariate normal distributions given by Patil and Boswell (1970). Their result leads to the equation

$$p(\xi_1) = \frac{1}{4} + \frac{\arcsin(\rho)}{2\pi} \quad , \tag{5.4}$$

where $\rho = \dfrac{-\sigma_2^2}{\sqrt{\left(\sigma_1^2 + \sigma_2^2\right)\left(\sigma_2^2 + \sigma_3^2\right)}}$. Although it may not at first be intuitively clear , in this example $P(X_3 < X_1 < X_2)$ and $P(X_2 < X_1 < X_3)$ both are equal and acts as $p_{\max}$. This can be explained by the fact that a normal distribution with a larger variance will result in more large and small observations than a normal distribution with the same mean and a smaller variance. With increasing s, $p(\xi_1)$ decreases. This case is presented in detail in Table 5.1.3, Plot 5.1.3(a) and Plot 5.1.3(b). The two plots are for $p(\xi_1)$ and $p_{\max}$, respectively. Again, here we assume X_1 is distributed as a standard normal. The next result is related to this one.

Theorem 5.3: Consider, $X_1 \sim N(0,1)$, $X_2 \sim N(0,s)$ and $X_3 \sim N(0,ls)$, $l > 2$ independent random variables.. Then, $p(\xi_1)$ is maximum at $s = l - 2$.

Proof: Using (5.4), if $X_1 \sim N(0,1)$, $X_2 \sim N(0,s)$ and $X_3 \sim N(0,ls)$,

$$\rho = \frac{-(1+s)}{\sqrt{(2+s)(2+(l+1)s)}}. \tag{5.5}$$

Note that, $\rho < 0$.

So, evidently, maximizing $p(\xi_1)$ is equivalent to minimizing (5.3).

To that end, differentiating ρ with respect to s, we have,

$$\frac{\partial \rho}{\partial s} = \frac{-\sqrt{(2+s)(2+(l+1)s)} + ((1+s)((l+2)+(l+1)s))/\sqrt{(2+s)(2+(l+1)s)})}{(2+s)(2+(l+1)s)}$$

$$= \frac{(l-2)-s}{(2+s)(2+(l+1)s)\sqrt{(2+s)(2+(l+1)s)}} \tag{5.6}$$

Equating (5.6) to 0 and noting that $\frac{\partial \rho}{\partial s}$ goes from positive to negative as s increases past $(l-2)$, $p(\xi_1)$ is maximum at $s = l-2$.

Case 4

Assume that for $d > 0$, $X_1 \sim N(\mu,1)$, $X_2 \sim N(\mu+d,1+s)$ and $X_3 \sim N(\mu+2d,1+2s)$ are independent random variables. The behavior of $p_{\max}$ which equals $p(\xi_1)$, is plotted in detail in Plot 5.1.4. As evident from the earlier examples, the parameter of interest in this case is most influenced by changes of location.

For larger values of K, with the help of the statistical software *R*, the precedence probabilities are found to behave in a somewhat similar fashion as when $K = 3$ if the changes happen either only in location or only in scale parameters. Namely, under scenarios for Case 1 and Case 3, $p(\xi_1)$ and $p_{\max}$ behave in a similar fashion. However, it gets more complicated if both means and variances are different among the distributions.

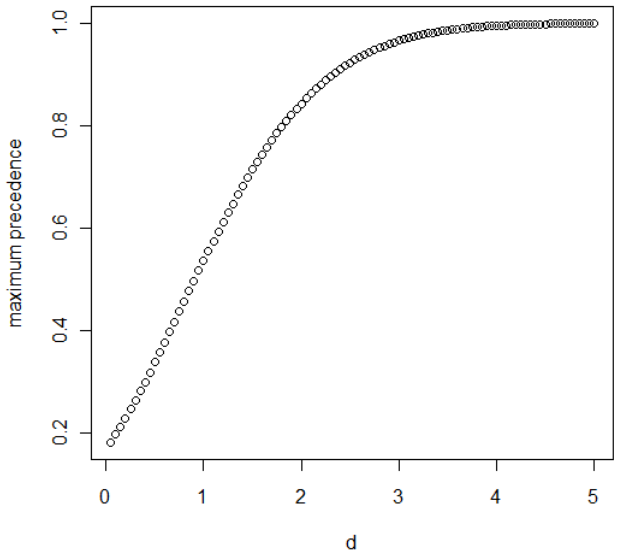
A few of the chosen examples are listed in Table 5.1.4 where both $p(\xi_1)$ and $p_{\max}$ are listed for different combinations of underlying normal distributions. As usual, we assume X_1 is distributed as a standard normal.

Table 5.1.1: Maximum Precedence for $X_1 \sim N(0,1)$, $X_2 \sim N(d,1)$ and $X_3 \sim N(2d,1)$

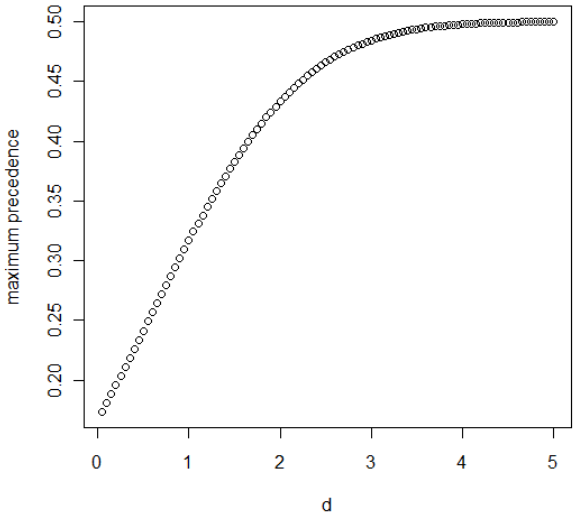
F_1	F_2	F_3	d	$p(\xi_1) = p_{\max}$
N(0,1)	N(0,1)	N(0,1)	0	0.1666667
	N(0.05,1)	N(0.1,1)	0.05	0.1811128
	N(0.1,1)	N(0.2,1)	0.10	0.1962275
	N(0.15,1)	N(0.3,1)	0.15	0.2119855
	N(0.2,1)	N(0.4,1)	0.20	0.2283567
	N(0.25,1)	N(0.5,1)	0.25	0.2453063
	N(0.3,1)	N(0.6,1)	0.30	0.2627954
	N(0.35,1)	N(0.7,1)	0.35	0.2807809
	N(0.4,1)	N(0.8,1)	0.40	0.2992165
	N(0.5,1)	N(1,1)	0.50	0.3372375
	N(0.6,1)	N(1.2,1)	0.60	0.3764365
	N(0.7,1)	N(1.4,1)	0.70	0.4163691
	N(0.8,1)	N(1.6,1)	0.80	0.4565863
	N(0.9,1)	N(1.8,1)	0.90	0.4966508
	N(1,1)	N(2,1)	1.00	0.5361516
	N(1.2,1)	N(2.4,1)	1.20	0.6120176
	N(1.4,1)	N(2.8,1)	1.40	0.6817908
	N(1.6,1)	N(3.2,1)	1.60	0.7439262
	N(1.8,1)	N(3.6,1)	1.80	0.7976885
	N(2,1)	N(4,1)	2.00	0.8430121
	N(2.5,1)	N(5,1)	2.50	0.9229231
	N(3,1)	N(6,1)	3.00	0.9661062
	N(3.5,1)	N(7,1)	3.50	0.9866717
	N(4,1)	N(8,1)	4.00	0.9953223
	N(4.5,1)	N(9,1)	4.50	0.9985373
	N(5,1)	N(10,1)	5.00	0.9995930

Table 5.1.2: Maximum Precedence for $X_1 \sim N(0,1)$, $X_2 \sim N(0,1)$ and $X_3 \sim N(d,1)$

F_1	F_2	F_3	d	$p(\xi_1) = p_{\max}$
N(0,1)	N(0,1)	N(0,1)	0	0.1666667
		N(0.05,1)	0.05	0.1737750
		N(0.1,1)	0.10	0.1809891
		N(0.15,1)	0.15	0.1882994
		N(0.2,1)	0.20	0.1956962
		N(0.25,1)	0.25	0.2031691
		N(0.3,1)	0.30	0.2107075
		N(0.35,1)	0.35	0.2183006
		N(0.4,1)	0.40	0.2259374
		N(0.5,1)	0.50	0.2412964
		N(0.6,1)	0.60	0.2566933
		N(0.7,1)	0.70	0.2720363
		N(0.8,1)	0.80	0.2872347
		N(0.9,1)	0.90	0.3022007
		N(1,1)	1.00	0.3168510
		N(1.2,1)	1.20	0.3449002
		N(1.4,1)	1.40	0.3708489
		N(1.6,1)	1.60	0.3943010
		N(1.8,1)	1.80	0.4150111
		N(2,1)	2.00	0.4328836
		N(2.5,1)	2.50	0.4656826
		N(3,1)	3.00	0.4843977
		N(3.5,1)	3.50	0.4937048
		N(4,1)	4.00	0.4977483
		N(4.5,1)	4.50	0.4992863
		N(5,1)	5.00	0.4997996



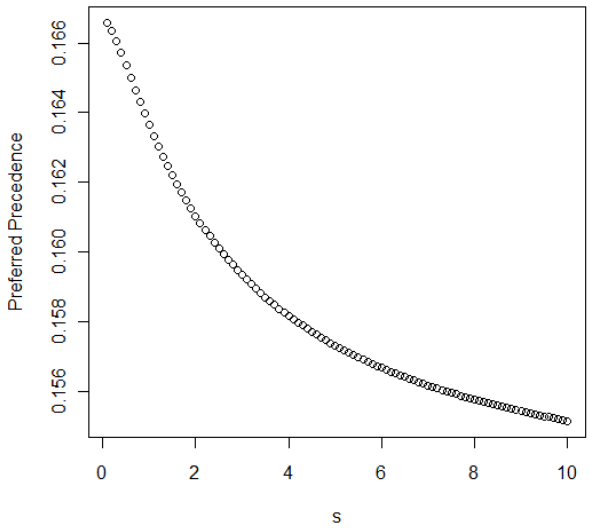
Plot 5.1.1: Maximum Precedence when $X_1 \sim N(0,1)$, $X_2 \sim N(d,1)$ and $X_3 \sim N(2d,1)$



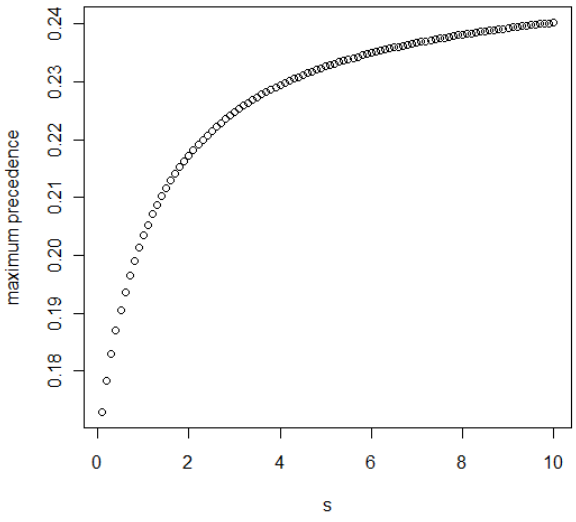
Plot 5.1.2: Maximum Precedence when $X_1 \sim N(0,1)$, $X_2 \sim N(0,1)$ and $X_3 \sim N(d,1)$

Table 5.1.3: Maximum and Preferred Precedence for $X_1 \sim N(0,1)$, $X_2 \sim N(0,1+s)$, $X_3 \sim N(0,1+2s)$

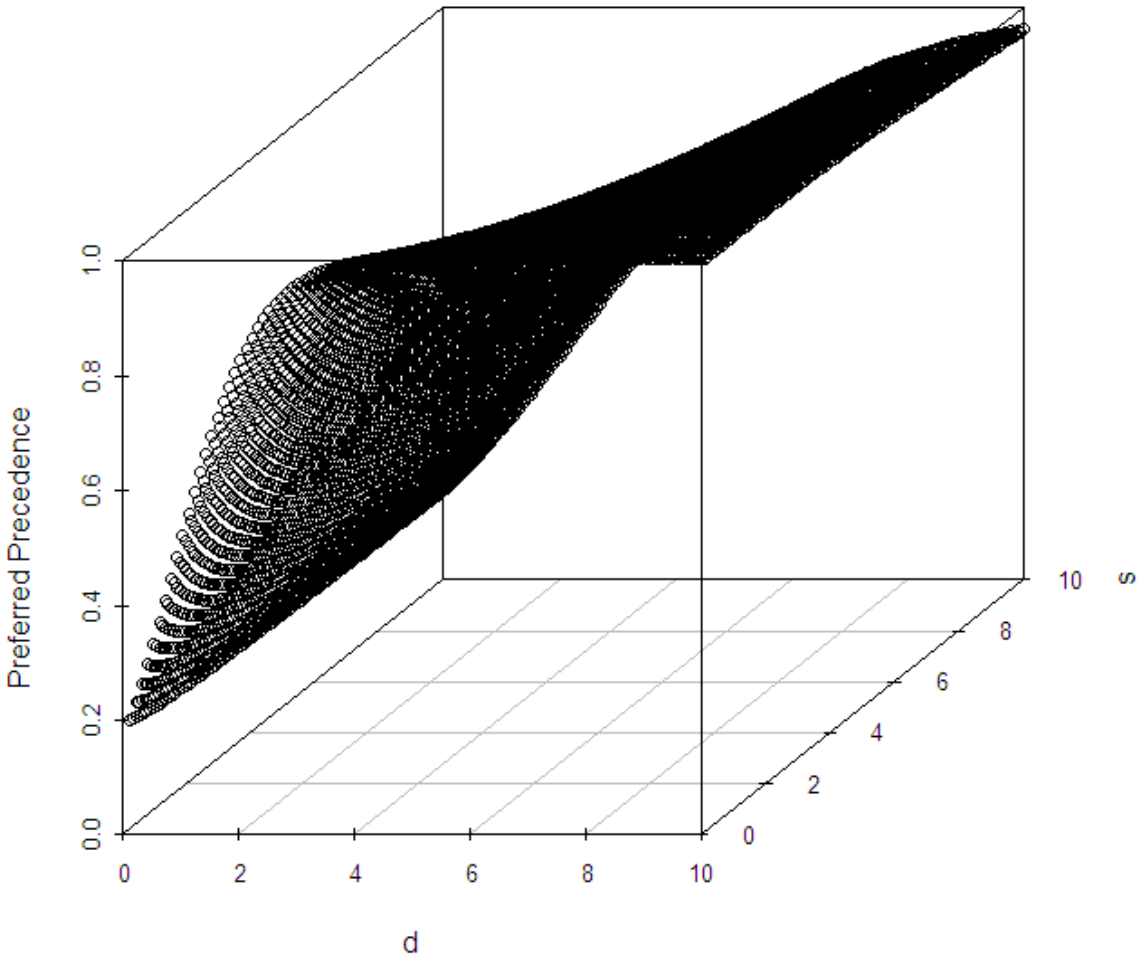
F_1	F_2	F_3	s	$p(\xi_1)$	$p_{\max}$
N(0,1)	N(0,1)	N(0,1)	0	0.1666667	0.1666667
	N(0,1.1)	N(0,1.2)	0.1	0.1665716	0.1729836
	N(0,1.2)	N(0,1.4)	0.2	0.1663458	0.1783399
	N(0,1.3)	N(0,1.6)	0.3	0.1660481	0.1829509
	N(0,1.4)	N(0,1.8)	0.4	0.1657128	0.1869699
	N(0,1.5)	N(0,2)	0.5	0.1653601	0.1905091
	N(0,1.6)	N(0,2.2)	0.6	0.1650025	0.1936532
	N(0,1.7)	N(0,2.4)	0.7	0.1646476	0.1964673
	N(0,1.8)	N(0,2.6)	0.8	0.1642999	0.1990027
	N(0,2)	N(0,3)	1	0.1636363	0.2033926
	N(0,2.2)	N(0,3.4)	1.2	0.1630216	0.2070662
	N(0,2.4)	N(0,3.8)	1.4	0.1624571	0.2101894
	N(0,2.6)	N(0,4.2)	1.6	0.1619403	0.2128797
	N(0,2.8)	N(0,4.6)	1.8	0.1614673	0.2152227
	N(0,3)	N(0,5)	2	0.1610340	0.2172827
	N(0,3.4)	N(0,5.8)	2.4	0.1602706	0.2207390
	N(0,3.8)	N(0,6.6)	2.8	0.1596218	0.2235274
	N(0,4.2)	N(0,7.4)	3.2	0.1590650	0.2258259
	N(0,4.6)	N(0,8.2)	3.6	0.1585828	0.2277542
	N(0,5)	N(0,9)	4	0.1581614	0.2293957
	N(0,6)	N(0,11)	5	0.1573102	0.2326001
	N(0,7)	N(0,13)	6	0.1566653	0.2349388
	N(0,8)	N(0,15)	7	0.1561604	0.2367217
	N(0,9)	N(0,17)	8	0.1557545	0.2381263
	N(0,10)	N(0,19)	9	0.1554212	0.2392616
	N(0,11)	N(0,21)	10	0.1551427	0.2401985



Plot 5.1.3(a): Preferred Precedence for $X_1 \sim N(0,1)$, $X_2 \sim N(0,1+s)$, $X_3 \sim N(0,1+2s)$



Plot 5.1.3(b): Maximum Precedence for $X_1 \sim N(0,1)$, $X_2 \sim N(0,1+s)$, $X_3 \sim N(0,1+2s)$



Plot 5.1.4: Preferred Precedence for $X_1 \sim N(0,1)$, $X_2 \sim N(d,1+s)$, $X_3 \sim N(2d,1+2s)$

Table 5.1.4: Maximum and Preferred Precedence for $X_1 \sim N(0,1)$, $X_2 \sim N(d_1,(1+s_1))$ and $X_3 \sim N(d_2,(1+s_2))$

F_1	F_2	F_3	$p(\xi_1)$	$p_{\max}$
N(0,1)	N(0,1)	N(0,1)	0.166667	0.166667
	N(1,1)	N(0,1)	0.126548	0.316851
	N(2,1)	N(0,1)	0.055583	0.432884
	N(2,1)	N(1,1)	0.1926	0.536152
	N(0,1)	N(1,2)	0.290151	0.290151
	N(0,1)	N(1,0.5)	0.339402	0.339402
	N(0,0.5)	N(1,0.5)	0.442493	0.442493
	N(1,0.5)	N(1,0.5)	0.373389	0.373389
	N(0,2)	N(1,1)	0.208362	0.3406
	N(0,2)	N(1,0.5)	0.205037	0.364364
	N(1,2)	N(1,0.5)	0.194307	0.395651
	N(1,2)	N(2,3)	0.341339	0.341339
	N(1,2)	N(2,2)	0.345068	0.345068
	N(1,2)	N(2,1)	0.354583	0.354583
	N(1,2)	N(2,0.5)	0.361739	0.361739
	N(1,1)	N(2,0.5)	0.579117	0.579117
	N(1,0.75)	N(2,0.5)	0.659266	0.659266
	N(2,1.5)	N(4,2)	0.659266	0.659266
	N(0,1)	N(3,1.25)	0.473685	0.473685
	N(1,1)	N(3,1.25)	0.660180	0.660180
	N(1,1.25)	N(3,1.25)	0.609918	0.609918
	N(1,1)	N(3,1)	0.684116	0.684116
	N(1,1)	N(4,1)	0.743529	0.743529
	N(0,1)	N(0.1,0.5)	0.15936	0.216916
	N(2,1)	N(6,1)	0.919012	0.919012
	N(3,1)	N(9,1)	0.983042	0.983042
	N(4,1)	N(16,1)	0.997661	0.997661
	N(0,1)	N(10,1)	0.5	0.5
	N(0,10)	N(0,100)	0.234216	0.249842
	N(0,0.01)	N(0,0.001)	0.248416	0.249984
	N(8,1)	N(16,1)	1	1

5.2 *The Exponential Distribution*

The Exponential distribution is an interesting situation since the same parameter can be thought of as both a location and a scale parameter.

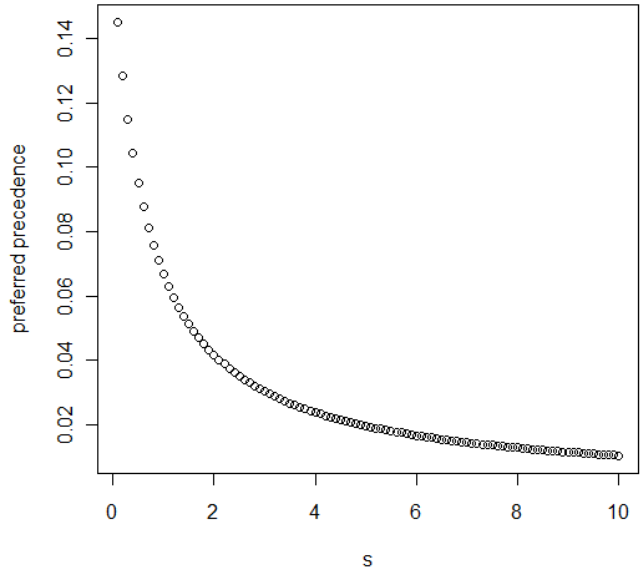
Specifically, let independently distributed $X_i \sim Exp(\lambda_i)$, $i = 1,2,3$, so that

$$E(X_i) = \frac{1}{\lambda_i}.$$

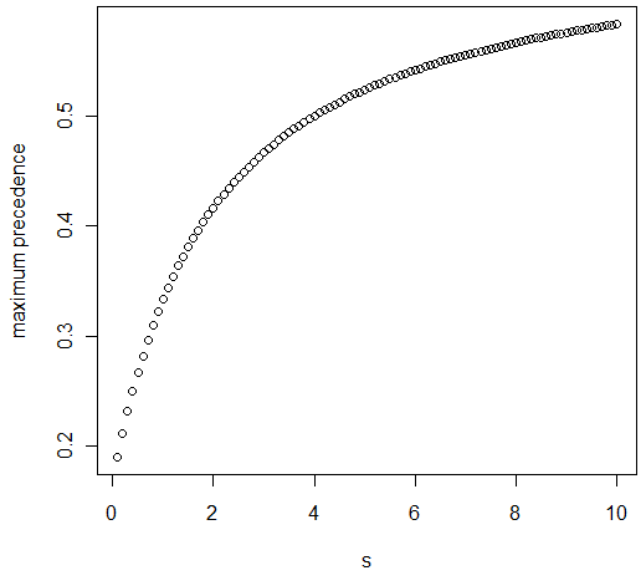
It then follows from a straightforward computation that

$$P(X_1 < X_2 < X_3) = \frac{\lambda_1 \lambda_2}{(\lambda_2 + \lambda_3)} \frac{1}{(\lambda_1 + \lambda_2 + \lambda_3)} \quad . \tag{5.7}$$

Without loss of generality, take $\lambda_1 = 1$ and set $\lambda_2 = 1 + s$ and $\lambda_3 = 1 + 2s$, with $s > 0$. Then , $p(\xi_1) = P(X_1 < X_2 < X_3) = \frac{1}{3(2+3s)}$ is a decreasing function of s. Table 5.2.1 provides a detailed summary. Plot 5.2.1 plots $p(\xi_1)$ against s, while Plot 5.2.2 plots the maximum precedence $p_{\max}$ against s. In this case, $P(X_3 < X_2 < X_1)$ is the maximum precedence probability, $p_{\max}$. Table 5.2.2 gives details for both $p(\xi_1)$ and $p_{\max}$ in a more general scenario, while keeping $X_i \sim Exp(\lambda_i)$, $i = 1,2,3$.



Plot 5.2.1: Preferred Precedence for $X_1 \sim Exp(1)$, $X_2 \sim Exp(1+s)$, $X_3 \sim Exp(1+2s)$



Plot 5.2.2: Maximum Precedence for $X_1 \sim Exp(1)$, $X_2 \sim Exp(1+s)$, $X_3 \sim Exp(1+2s)$

Table 5.2.1: Maximum and Preferred Precedence for $X_1 \sim Exp(1)$, $X_2 \sim Exp(1+s)$, $X_3 \sim Exp(1+2s)$

F_1	F_2	F_3	s	$p(\xi_1)$	$p_{\max}$
Exp(1)	Exp(1)	Exp(1)	0	0.16666667	0.1666667
	Exp(1.1)	Exp(1.2)	0.1	0.14492754	0.1904762
	Exp(1.2)	Exp(1.4)	0.2	0.12820513	0.2121212
	Exp(1.3)	Exp(1.6)	0.3	0.11494253	0.2318841
	Exp(1.4)	Exp(1.8)	0.4	0.10416667	0.2500000
	Exp(1.5)	Exp(2)	0.5	0.09523810	0.2666667
	Exp(1.6)	Exp(2.2)	0.6	0.08771930	0.2820513
	Exp(1.7)	Exp(2.4)	0.7	0.08130081	0.2962963
	Exp(1.8)	Exp(2.6)	0.8	0.07575758	0.3095238
	Exp(2)	Exp(3)	1	0.06666667	0.3333333
	Exp(2.2)	Exp(3.4)	1.2	0.05952381	0.3541667
	Exp(2.4)	Exp(3.8)	1.4	0.05376344	0.3725490
	Exp(2.6)	Exp(4.2)	1.6	0.04901961	0.3888889
	Exp(2.8)	Exp(4.6)	1.8	0.04504505	0.4035088
	Exp(3)	Exp(5)	2	0.04166667	0.4166667
	Exp(3.4)	Exp(5.8)	2.4	0.03623188	0.4393939
	Exp(3.8)	Exp(6.6)	2.8	0.03205128	0.4583333
	Exp(4.2)	Exp(7.4)	3.2	0.02873563	0.4743590
	Exp(4.6)	Exp(8.2)	3.6	0.02604167	0.4880952
	Exp(5)	Exp(9)	4	0.02380952	0.5000000
	Exp(6)	Exp(11)	5	0.01960784	0.5238095
	Exp(7)	Exp(13)	6	0.01666667	0.5416667
	Exp(8)	Exp(15)	7	0.01449275	0.5555556
	Exp(9)	Exp(17)	8	0.01282051	0.5666667
	Exp(10)	Exp(19)	9	0.01149425	0.5757576
	Exp(11)	Exp(21)	10	0.01041667	0.5833333

Table 5.2.2: Maximum and Preferred Precedence for $X_1 \sim Exp(\lambda_1)$, $X_2 \sim Exp(\lambda_2)$, $X_3 \sim Exp(\lambda_3)$

F_1	F_2	F_3	$p(\xi_1)$	$p_{\max}$
Exp(1)	Exp(1)	Exp(1)	0.166667	0.166667
	Exp(1)	Exp(2)	0.083333	0.25
	Exp(1)	Exp(0.5)	0.266667	0.266667
	Exp(0.25)	Exp(0.5)	0.095238	0.380952
	Exp(1)	Exp(4)	0.033333	0.333333
	Exp(3)	Exp(4)	0.075	0.375
	Exp(3)	Exp(6)	0.0428571	0.45
	Exp(3)	Exp(9)	0.0230769	0.519231
	Exp(0.33)	Exp(0.66)	0.099897	0.335008
	Exp(1)	Exp(0.66)	0.224271	0.224271
	Exp(1)	Exp(0.33)	0.322695	0.322695
	Exp(1)	Exp(0.25)	0.355556	0.355556
	Exp(1)	Exp(0.1)	0.4329	0.4329
	Exp(1)	Exp(0.01)	0.492587	0.492587
	Exp(2)	Exp(0.01)	0.657873	0.657873
	Exp(4)	Exp(0.01)	0.790498	0.790498
	Exp(10)	Exp(0.01)	0.899272	0.899272
	Exp(4)	Exp(0.25)	0.609524	0.609524
	Exp(4)	Exp(16)	0.011204	0.609524
	Exp(4)	Exp(8)	0.034188	0.492308
	Exp(6)	Exp(11)	0.044444	0.457143
	Exp(0.95)	Exp(1.05)	0.154472	0.179487
	Exp(0.75)	Exp(1.25)	0.111111	0.238095
	Exp(0.05)	Exp(25)	0.000007	0.913992
	Exp(1.5)	Exp(0.5)	0.333333	0.333333

5.3 *The Logistic Distribution*

A study of the logistic distribution helps us understand the behavior of precedence probabilities for symmetric distributions having heavier tails than the normal distribution. If X is a random variable that follows a logistic distribution with location parameter μ and scale parameter σ, its distribution function F is given by

$$F(x) = 1 - \frac{1}{\exp((x-\mu)/\sigma)}. \tag{5.8}$$

Unfortunately, calculating exact precedence probabilities, particularly for large K is not simple. The aim here is to check the behavior of precedence probabilities. So, although an exact solution would have been more effective; I opted for using simulations to obtain approximate precedence probabilities. For this purpose, I used the software *R*. Thirty (30) observations are generated at random from each of the distributions. Then, the precedence probabilities are estimated by counting the total number of occurrences of a particular precedence and dividing that by the total number of possible occurrences of that precedence. For example, let, $K = 3$. Then, obtain a sample of size 1 from each of the 30 sets of observations corresponding to different distributions. Clearly, there are $30 \times 30 \times 30 = 27000$ such triplets possible out of the given observations. To estimate $p(\xi_1)$, we count the no. of triplets in which the observation from F_3 is the largest while the observation from F_1 is the smallest;

and then divide that number by 27000. Like the preferred precedence, all other precedence probabilities are estimated in this way. This procedure is then repeated 1000 times and the average of the estimated precedence probabilities are used as the final estimate of these precedence probabilities and the maximum precedence is estimated as the maximum of the final estimates of precedence probabilities.

The estimated precedence probabilities are summarized in detail in Table 5.3.1, given below, for a few selected scenarios where I took $X_1 \sim Logis(0,1)$, $X_2 \sim Logis(d_1, s_1)$ and $X_3 \sim Logis(d_2, s_2)$ and used the notation L to denote a logistic distribution. It appears from these data that the maximum precedence in the logistic case is smaller than the maximum precedence in the normal case for the same location and scale parameters. That is, however, expected, since the logistic distribution boasts of a heavier tail.

5.4 *The Generalized Extreme Value Distribution*

Our intention here is to check the behavior of precedence probabilities for an asymmetric distribution. The generalized extreme value distribution seemed to be a good candidate, as it deals with the asymmetry via the introduction of a

third parameter, called a shape parameter into the probability density function model.

If X is a random variable that follows a generalized extreme value c distribution with location parameter μ, scale parameter σ and shape parameter ψ the distribution function F of the said variable for $\psi > 0$ is given

$$F(x) = \exp\left(-\left(1+\psi\left(\frac{x-\mu}{\sigma} \right) \right)^{-1/\psi} \right) \text{ , if } 1+\psi\left(\frac{x-\mu}{\sigma} \right) > 0 \tag{5.9}$$

It is to be noted that the support for this distribution depends on the shape parameter. For $\psi = 0$, the support is the whole real line. The mean of the distribution is ∞ if $\psi \geq 1$. Since obtaining exact precedence probabilities is difficult, I used a similar simulation approach to the one used in 5.3. Again using the software package *R*, thirty (30) observations were generated at random from each of the distributions. Then, the precedence probabilities were estimated by counting the total number of occurrences of a specified precedence and then dividing that by the total number of possible occurrences of that precedence. This process is then repeated 1000 times and the average of the estimated precedence probabilities are used as the final estimates of these precedence probabilities and the maximum precedence is estimated as the maximum of the final estimates of precedence probabilities. So, instead of actual precedence probabilities, we used these estimates to check their behavior. My results are summarized in detail in Table 5.3.1 for few selected

cases. Specifically, I took $X_1 \sim GEV(0, scale = 1, shape = \psi)$, $X_2 \sim GEV(d_1, s_1, \psi_1)$ and $X_3 \sim GEV(d_2, s_2, \psi_2)$, where GEV denotes a generalized extreme value distribution.

Here, as the shape parameter gets larger, we observe that the maximum precedence probability gets smaller for distributions with the same set of location and scale parameters. And as usual, the maximum precedence gets larger if distributions with the larger location have the smaller scale parameters.

Table 5.3.1: Maximum and Preferred Precedence for $X_1 \sim Logis(0,1)$, $X_2 \sim Logis(d_1, s_1)$ and $X_3 \sim Logis(d_2, s_2)$

F_1	F_2	F_3	$\hat{p}(\xi_1)$	$\hat{p}_{\max}$
L(0,1)	L(0,1)	L(0,1)	0.1681	0.1681
	L(1,1)	L(2,1)	0.3768	0.3768
	L(0,2)	L(0,3)	0.1701	0.2268
	L(1,2)	L(2,3)	0.2684	0.2684
	L(2,2)	L(4,3)	0.3832	0.3832
	L(2,3)	L(4,3)	0.3028	0.3028
	L(2,1)	L(4,3)	0.4994	0.4994
	L(2,0.5)	L(4,3)	0.5618	0.5618
	L(0,1)	L(2,1)	0.3454	0.3454
	L(0,1)	L(2,2)	0.3121	0.3121
	L(0,1)	L(2,0.5)	0.3737	0.3737
	L(3,2)	L(6,3)	0.4995	0.4995
	L(3,2)	L(6,4)	0.4786	0.4786
	L(1,2)	L(2,1)	0.2448	0.2812
	L(1,2)	L(2,0.5)	0.2393	0.3240
	L(1,1)	L(2,0.5)	0.3927	0.3927
	L(1,0.75)	L(2,0.5)	0.4587	0.4587
	L(2,1.5)	L(4,2)	0.4583	0.4583
	L(0,1)	L(3,1.25)	0.3974	0.3974
	L(1,1)	L(3,1.25)	0.4677	0.4677
	L(1,1.25)	L(3,1.25)	0.4248	0.4248
	L(1,1)	L(3,1)	0.4843	0.4843
	L(1,1)	L(4,1)	0.5635	0.5635
	L(2,1)	L(6,1)	0.7443	0.7443
	L(3,1)	L(9,1)	0.8801	0.8801
	L(4,1)	L(16,1)	0.9470	0.9470
	L(0,1)	L(10,1)	0.5071	0.5071
	L(0,10)	L(0,100)	0.2389	0.2514
	L(0,0.01)	L(0,0.001)	0.2475	0.2577
	L(8,1)	L(16,1)	0.9956	0.9956
	L(15,1)	L(30,1)	1	1

Table 5.4.1: Maximum and Preferred Precedence $X_1 \sim GEV(0,1,\psi)$, $X_2 \sim GEV(d_1,s_1,\psi_1)$ and $X_3 \sim GEV(d_2,s_2,\psi_2)$

F_1	F_2	F_3	$\hat{p}(\xi_1)$	$\hat{p}_{\max}$
G(0,1,0)	G(1,1,0)	G(2,1,0)	0.4909	0.4909
	G(0,2,0)	G(0,3,0)	0.2089	0.2317
	G(1,2,0)	G(2,3,0)	0.3778	0.3778
	G(2,2,0)	G(4,3,0)	0.5665	0.5665
	G(2,1,0)	G(4,3,0)	0.6881	0.6881
	G(2,0.5,0)	G(4,3,0)	0.7348	0.7348
	G(2,0.5,0)	G(4,0.5,0)	0.8766	0.8766
	G(0,1,0)	G(2,1,0)	0.3990	0.3990
G(0,1,0.5)	G(1,1,0.5)	G(2,1,0.5)	0.4460	0.4460
	G(1,1.5,0)	G(2,2,-.5)	0.3580	0.3580
	G(1,2,0)	G(2,1.5,-.5)	0.2878	0.2878
	G(1,2,-.5)	G(2,1.5,0)	0.3502	0.3502
	G(1,1.5,-.5)	G(2,2,0)	0.4074	0.4074
G(0,1,0)	G(1,1.5,-.5)	G(2,2,.5)	0.4584	0.4584
	G(1,2,-.5)	G(2,1.5,.5)	0.4053	0.4053
	G(1,2,.5)	G(2,1.5,-.5)	0.2917	0.3535
	G(1,1.5,.5)	G(2,2,-.5)	0.3582	0.3582
G(0,1,-.5)	G(1,1,-.5)	G(2,1,-.5)	0.5784	0.5784
	G(2,1,-.5)	G(4,1,-.5)	0.8748	0.8748
	G(1,1.5,0)	G(2,2,0.5)	0.4774	0.4774
	G(1,1.5,.5)	G(2,2,0)	0.4179	0.4179
	G(1,2,0)	G(2,1.5,.5)	0.4084	0.4084
	G(1,2,.5)	G(2,1.5,0)	0.3500	0.3734
	G(0,1,0)	G(0,1,.5)	0.2148	0.2148
G(0,1,0)	G(0,1,-1)	G(0,1,1)	0.2129	0.2534
G(0,1,0)	G(0,1,-.2)	G(0,1,.2)	0.1864	0.1869
G(0,1,.2)	G(1,1,.2)	G(2,1,.2)	0.4692	0.4692
G(0,1,-.2)	G(1,1,-.2)	G(2,1,-.2)	0.5204	0.5204
G(0,1,-1)	G(1,1,-1)	G(2,1,-1)	0.6419	0.6419
G(0,1,1)	G(1,1,1)	G(2,1,1)	0.4197	0.4197
G(0,1,0)	G(0,10,0)	G(0,20,0)	1	1

From this study, we note the following common behavior of precedence probabilities irrespective of the type of parametric distribution considered above. Of course, these are just a very small set of possibilities and are not intended as an exhaustive study. In the cases of two-parameter families, $p_{\max}$ gets larger if the distributions with higher location parameters tend to have smaller scale parameters. $p_{\max} = p(\xi_1)$ if the location parameter of distribution 1 is smaller than the location parameter of distribution 2 which in turn is smaller than the location parameter of distribution 3. Unless the differences between these parameters are small and the distribution with the largest location parameter has the smallest scale parameter, $p_{\max}$ tends to get larger as the difference between distributions increases.

Based on these observations, we introduce the concept of distance-index measures among distributions in the next chapter.

Chapter 6: New Index Measures for Separation and Similarity among Distributions

Rejection of the null hypothesis that the K distributions $\{F_i; i = 1, 2, ..., K\}$ are identical naturally leads to the important issue of measuring and describing just how different the distributions are. Most existing work in this area is fully parametric, based on functionals such as means, variances and on densities and applies only to the case K = 2. My use of precedence probabilities to address this issue is a new approach that leads to non-parametric measures that are invariant with respect to monotone increasing transformations, have observable consequences and apply to the general case $K \geq 2$.

Defining a useful, intuitively appealing *dissimilarity* or *distance* measure among three or more distributions is a difficult problem to which there is clearly no unique, best solution. Bioequivalence or biosimilarity is an application where such a distance measure is very important. *Biosimilar* is a term used to describe an officially-approved, later version of an already existing biopharmaceutical product made by a different producer following patent expiration of the original product. The United States Food and Drug Administration (FDA) in 2003 has defined bioequivalence as, "The absence of a significant difference in the rate and extent to which the active ingredient or active moiety in pharmaceutical equivalents or pharmaceutical alternatives

becomes available at the site of drug action when administered at the same molar dose under similar conditions in an appropriately designed study." If two pharmaceutical products are bioequivalent, essentially they are the same for all intents and purposes. For example, a new drug which is bioequivalent to an existing drug can generate a market share for itself with a smaller market price. To prove bioequivalence, some measures of distances are needed.

Popular measures of pair-wise dissimilarity require exact or approximate specification of the distribution functions or their densities and increasing values of measures correspond to increasing differences among the distributions. Best known among these is the Kullback-Liebler divergence. For any two continuous random variables with continuous density functions f_1 and f_2 and distribution functions F_1 and F_2, the Kullback-Liebler divergence is defined by

$$\mathrm{D}_{KL}(F_1 \,||\, F_2) = \int_{-\infty}^{\infty} f_1(x) \log\left(\frac{f_1(x)}{f_2(x)}\right) dx \qquad (6.1)$$

The Hellinger distance between two distributions, first introduced by Cramer (1946) in terms of the Hellinger Integral is defined as

$$\mathrm{D}_H(F_1, F_2) = \sqrt{\int_{-\infty}^{\infty} \left(\sqrt{f_1(x)} - \sqrt{f_2(x)}\right)^2 dx} \quad . \qquad (6.2)$$

The Hellinger distance can also expressed in terms of what is referred to as *affinity*, developed by Bhattacharya (1943) and defined by

$$\mathrm{D}_B(F_1, F_2) = -\ln\left(\int_{-\infty}^{\infty} \left(\sqrt{f_2(x) f_1(x)} \right) dx \right) \tag{6.3}$$

The Cramer-Von Mises criterion (Anderson, 1962) and Kolmogorov-Smirnov two sample test statistics can also be used to define distances between two distributions. Specifically, the population analogues of the two-sample Kolmogorov-Smirnov test statistic and of the Cramer-Von Mises criterion are given, respectively, by

$$\mathrm{D}_{KS}(F_1, F_2) = Sup_x \mid F_1(x) - F_2(x) \mid \tag{6.4}$$

$$\mathrm{D}_{CVM}(F_1, F_2) = \int \left(F_1(x) - F_2(x) \right)^2 dF_{1+2}(x), \tag{6.5}$$

where F_{1+2} is a distribution function obtained as

$$F_{1+2}(x) = \frac{1}{2}\left[F_1(x) + F_2(x) \right]. \tag{6.6}$$

A promising new measure of the difference between two distributions was developed by Szekely in 1989. This measure is known as the Energy distance measure and is defined by

$$D(F_1, F_2) = 2E \parallel X_1 - X_2 \parallel - E \parallel X_1 - X_1^* \parallel - E \parallel X_2 - X_2^* \parallel \tag{6.7}$$

where X_1 and X_1^* are independently and identically distributed (i.i.d) as F_1, X_2 and X_2^* are i.i.d F_2; $\parallel . \parallel$ is a vector norm and E denotes the expected value. This distance was recently used by Rizzo and Szekely (2010) to develop a new

type of nonparametric analysis of variance, which has been discussed in some detail in the chapters 4 and 10.

Another measure of distance which is gaining popularity is the so-called Earth-Mover's Distance based on an idea by Monge (1781) and first applied by Peleg, Werman and Rom (1989).

The effect-size parameter

$$\gamma = \sum_{j=1}^{K} (\mu_i - \bar{\mu}) / \sigma^2 \ , \tag{6.8}$$

given in Cohen (1988) has been widely used in the social sciences to describe difference among K distributions with means $\{\mu_i\}$ and common variance σ^2, where $\bar{\mu} = \sum \mu_i / K; i = 1, 2, .., K$.

Based on the precedence probability described earlier, we propose three new measures of similarity or separation between K distributions. Recall that there are $K!$ precedence probabilities $\{p(\xi_i); i = 1, 2, ..., K!\}$ among K distributions. In my opinion, differences among these distributions are well captured by these precedence probabilities. Recall that in the extreme cases, when all the distributions are equal, all the precedence probabilities equal $1/K!$ and when all the distributions are singular, $\{p(\xi_i); i = 1, 2, ..., K!\}$ all equal 1, their maximum value. In between these extremes, increasing variation and increasing magnitudes among the precedence probabilities correspond to increasing

differences among the distributions. Accordingly, I now introduce three summary measures of the precedence probabilities, defined as follows:

$$\gamma_1 = \max_i \{p(\xi_i); i = 1, 2, ..., k!\} \quad . \tag{6.9}$$

$$\gamma_2 = \sum_{i=1}^{K!} p(\xi_i)^2 \tag{6.10}$$

$$\gamma_3 = s.d.\{p(\xi_i); i = 1, 2, ..., k!\} \quad . \tag{6.11}$$

In (6.11), *s.d.* stands for standard deviation. Next, I present some properties of these summary measures in the form of a theorem.

Theorem 6.1: Let $\{p(\xi_i); i = 1, 2, ..., K!\}$ be the precedence probabilities arising from K continuous distributions and γ_1, γ_2 and γ_3 be as defined in (6.9), (6.10) and (6.11). Then, among all such possible distributions,

(i) $\max.\{\gamma_1\} = 1$

(ii) $\min.\{\gamma_1\} = 1/K!$

(iii) $\max.\{\gamma_2\} = 1$

(iv) $\min.\{\gamma_2\} = 1/K!$

(v) $\max.\{\gamma_3\} = \sqrt{\left(\frac{1}{K!}\right)}$

(vi) $\min.\{\gamma_3\} = 0$

Moreover, the maximum values for γ_1, γ_2 and γ_3 occur when one of the precedence probabilities is 1. Their minimum values occur when all of the precedence probabilities are equal.

Proof: Since, $\{p(\xi_i); i = 1,2,...,k!\}$ is a set of precedence probabilities,

$$\max_i \{p(\xi_i); i = 1,2,...,k!\} = \max.\{\gamma_1\} = 1, \tag{6.12}$$

which occurs when the distributions are mutually singular.

Since the precedence probabilities sum to 1,

$$mean(p(\xi_i); i = 1,2,...,K!) = \frac{1}{K!}. \tag{6.13}$$

Maximizing $s.d.(p(\xi_i); i = 1,2,...,K!)$ is equivalent to maximizing $\sum_{i=1}^{K!} p(\xi_i)^2$. Since $0 < p(\xi_i) < 1, i = 1,2,...,K!$, $p(\xi_i)^2 < p(\xi_i)$. Hence,

$$\sum_{i=1}^{K!} p(\xi_i)^2 \leq \sum_{i=1}^{K!} p(\xi_i) = 1. \tag{6.14}$$

If one of the precedence probabilities is 1, the inequality in (6.12) is an equality and

$$\max\{\gamma_2\} = 1. \tag{6.15}$$

Furthermore, when one of the precedence probabilities is 1,

$$s.d.(p(\xi_i); i = 1,2,...,K!) = \sqrt{\left(1-\frac{1}{K!}\right)^2 \frac{1}{K!-1} + \frac{1}{K!-1}\frac{(K!-1)}{K!^2}} = \sqrt{\left(\frac{1}{K!}\right)}.$$

That leads to,

$$\max.\{\gamma_3\} = \sqrt{\left(\frac{1}{K!}\right)} \quad . \tag{6.16}$$

It is to be noted that the mean of any set of observations cannot be greater than their maximum and that the mean and maximum are equal if and only if all the observations are identical. Specifically,

$$\max_i \{p(\xi_i); i = 1,2,...,k!\} \geq mean\{p(\xi_i); i = 1,2,...,K!\}.$$

Equality in the expression above occurs if and only if $p(\xi_i) = 1/K! \ \forall i = 1,2,...,K!$.

So, clearly,

$$\min.\{\gamma_1\} = 1/K!. \tag{6.17}$$

Also, since γ_3 is a standard deviation,

$$\min.\{s.d.(p(\xi_i); i = 1,2,...,K!)\} = \min.\{\gamma_3\} = 0 \tag{6.18}$$

which occurs if and only if $p(\xi_i) = 1/K! \forall i = 1,2,...,K!$.

Clearly,

$$\min.\{\gamma_2\} = 0, \tag{6.19}$$

and this occurs if and only if the underlying distributions are identical.

This completes the proof.

To gain some insight into the results presented in Theorem 6.1, consider the case when one of the precedence probabilities is 1. Without loss of generality, let $p(\xi_1) = P(X_1 < X_2 < \ldots < X_K) = 1$, which occurs when all observations from distribution function F_K would be larger than all observations from distribution function F_{K-1} and so on, with observations from distribution function F_1 being smallest among all. So, in this case there is a clear sense of ordering among distributions. This scenario leads to the maximum in each of the three proposed index measures. On the other hand, when all the distributions are identical , all the of precedence probabilities are equal, which in turn leads to all three proposed index measures equaling their minimum possible values.

Keeping the results obtained from Theorem 6.1 in mind, we introduce below standardized versions of (6.9) - (6.11). All three standardized measures range from 0 to 1, with a larger measure denoting very different distributions and a smaller measure denoting very slight differences among the distributions. A measure of 1 occurs only when the distributions are mutually singular and a measure of 0 occurs only when all the distributions are identical.

Here are the three proposed "standardized" measures γ_m, γ_{ss} and γ_{sd}:

$$\gamma_m = \frac{\gamma_1 - \frac{1}{K!}}{1 - \frac{1}{K!}}, \tag{6.20}$$

$$\gamma_{ss} = \frac{\gamma_2 - \frac{1}{K!}}{1 - \frac{1}{K!}}, \tag{6.21}$$

$$\gamma_{sd} = \frac{\gamma_3}{\sqrt{\left(\frac{1}{K!}\right)}}. \tag{6.22}$$

We have the following relationships between these indices.

Theorem 6.2: When, $K = 2$; $\gamma_m = \gamma_{sd}$.

Proof: Let, without any loss of generality, $\gamma_1 = p(\xi_1)$. Then

$$\gamma_m = 2(p(\xi_1) - 0.5). \tag{6.23}$$

On the other hand,

$$\gamma_3 = \sqrt{2(p(\xi_1) - 0.5)^2} \Rightarrow \gamma_{sd} = 2(p(\xi_1) - 0.5) = \gamma_m. \tag{6.24}$$

This completes the proof.

Theorem 6.3: $\gamma_{ss} = \gamma_{sd}^2$.

Proof: Note that,

$$\gamma_{sd}^2 = \frac{\sum_{i=1}^{K!}\left(p(\xi_i) - \left(\frac{1}{K!}\right)\right)^2}{\left(K!-1\right)} K!$$

$$= \frac{\sum_{i=1}^{K!} p(\xi_i)^2 - \frac{1}{K!}}{\left(K!-1\right)} K!$$

$$= \frac{\sum_{i=1}^{K!} p(\xi_i)^2 - \frac{1}{K!}}{\left(1-\frac{1}{K!}\right)} = \gamma_{ss} . \qquad (6.25)$$

Since, γ_{ss} and γ_{sd} are related in such a unique way, it is sufficient to study the behavior of any one of them. Depending on the ease of estimation, we will be able to propose a better index among these two measures.

In the next chapter we will explore the behavior of these distance indices under a variety of conditions.

Chapter 7: Behavior of Distance Index Measures

In this chapter, we are going to discuss the behavior of the distance measures among K distributions proposed in the previous chapter in (6.20)-(6.22). I am not going to compare them with any existing methods mostly because there are so many ways in which a "distance" can be defined. However, as specified in the earlier chapter, all of these measures attain their maximum values when all the distributions are mutually singular and all these measures attain their minimum values when all the distributions are identical.

I will start with the behavior of these measures when the underlying distributions are normal. I used the software *R* to calculate these indexes precisely with the help of the "mnormt" routine available. I used equation (5.3) to calculate which equates the precedence probability to the distribution function of a multivariate normal distribution with a specified correlation structure.

Case 1: $K = 2$ and both underlying distributions are normal

This is the simplest scenario for measuring the distance among distributions. For example, when the difference is only in the means of the distributions, a simple distance measure applied to the means suffices.. However, the problem

gets trickier if the variances of the distributions are not equal, which is probably an under-studied but important practical case. When, $K = 2$ and both underlying distributions are normal, we recall that the precedence probabilities are be equal if the means of the distributions are exactly equal, even if the scales differ. Also, as mentioned in Theorems 6.2 and 6.3, all the index measures given there are related to each other. Among the results seen in Table 7.1, note that based on my index measures, the distance between two normal distributions appears to be larger if the distribution with larger mean has the smaller variance.

Case 2: $K = 3$ and all the underlying distributions are normal.

This is, naturally the next step in checking the behavior of these index measures. When, $K = 3$ and all the underlying distributions are normal, we recall that the precedence probabilities are not necessarily equal, even if the means of the distributions are exactly same. In fact, if the variances differ for all the distributions, there are 3 separate pairs of precedence probabilities which are different from each other. Also, as mentioned in Theorem 6.3, index measures γ_{ss} and γ_{sd} are related to each other.

The results are listed in detail in Table 7.2. The findings here are, otherwise, similar to the previous case study in Table 7.1. If the distributions are different in variances only, clearly they do not seem to be that much different from

distributions which differ in location. This observation is reflected in the fact that, the maximum precedence seems to be converging to 0.25 and that, in turn, is reflected in the index measures, which do not approach the maximum possible value of 1. Again, that might be seen as strength of these measures.

Case 3: $K = 3$ and all the underlying distributions are exponential

This is an interesting case, as in the exponential distribution a single parameter acts as both a location and scale parameter. This case was studied in detail in Table 7.3. The precedence probability and hence, in turn, the index measures seem to depend on the ratio of the parameters of the exponential distributions. The index measures seem to be higher if the ratio of the smallest parameter to the middlemost parameter is the same as the ratio of the middlemost parameter to the largest parameter; rather than any other arrangement of the 3 parameters.

Case 4: $K = 3$ and all the underlying distributions are logistic.

For logistic distributions, the tails are heavier than normal distributions. So, logistic distributions might be expected to be less "different" than normal distributions having the same location and scale parameters. Here also, the precedence probabilities are necessarily equal if the location parameters of the distributions are the same. As seen in case 2, if the variances differ for all the

distributions, there are 3 separate pairs of precedence probabilities which are different from each other. A few selected results are given in Table 7.4. The findings here are very similar to those in case 2.

Case 5: $K = 3$ and all the underlying distributions are generalized extreme value distributions.

A generalized extreme value distribution has three separate parameters, one of which is a shape parameter responsible for the asymmetry of the distribution. This case study seems to be very similar to case 2 and case 4 in terms of location and scale parameters. However, the new dimension of shape parameter and its interaction with other parameters in determining precedence probabilities is presented in Table 7.5. These distributions do not seem at all different if they differ only in the shape parameters. Typically, higher maximum precedence is observed if the distribution with smallest location parameter is also associated with smallest shape parameter. If the scale parameter is involved, typically higher maximum precedence is observed if the distribution with smallest location and shape parameter is also associated with the smallest scale parameter.

Table 7.1: Behavior of Distance Index Measures when $K = 2$ and distributions are normal

F_1	F_2	γ_m	γ_{ss}	γ_{sd}
N(0,1)	N(0,1)	0	0	0
N(0,1)	N(1,1)	0.5205	0.2709	0.5205
N(0,1)	N(2,1)	0.8427	0.7101	0.8427
N(0,1)	N(3,1)	0.9661	0.9334	0.9661
N(0,1)	N(4,1)	0.9923	0.9907	0.9923
N(0,1)	N(5,1)	0.9996	0.9992	0.9996
N(0,1)	N(6,1)	1.0000	1.0000	1.0000
N(0,1)	N(1,2)	0.3453	0.1192	0.3453
N(0,1)	N(1,0.5)	0.6289	0.3955	0.6289
N(0,1)	N(1,4)	0.1916	0.0367	0.1916
N(0,1)	N(1,0.25)	0.6797	0.4620	0.6797
N(0,1)	N(2,2)	0.6289	0.3955	0.6289
N(0,1)	N(2,0.5)	0.9264	0.8581	0.9264
N(0,1)	N(2,4)	0.3724	0.1387	0.3724
N(0,1)	N(2,0.25)	0.9531	0.9085	0.9531
N(0,1)	N(3,2)	0.8203	0.6729	0.8203
N(0,1)	N(3,0.5)	0.9927	0.9854	0.9927
N(0,1)	N(3,4)	0.5331	0.2842	0.5331
N(0,1)	N(3,0.25)	0.9971	0.9943	0.9971

Table 7.2: Behavior of Distance Index Measures when $X_1 \sim N(0,1)$, $X_2 \sim N(d_1,(1+s_1))$ and $X_3 \sim N(d_2,(1+s_2))$

F_1	F_2	F_3	$p_{\max}$	γ_m	γ_{ss}	γ_{sd}
N(0,1)	N(0,1)	N(0,1)	0.166667	0	0	0
	N(1,1)	N(0,1)	0.316851	0.1802	0.0871	0.2951
	N(2,1)	N(0,1)	0.432884	0.3195	0.2575	0.5074
	N(2,1)	N(1,1)	0.536152	0.4434	0.2366	0.4865
	N(0,1)	N(1,2)	0.290151	0.1482	0.0556	0.2359
	N(0,1)	N(1,0.5)	0.339402	0.2072	0.1221	0.3495
	N(0,0.5)	N(1,0.5)	0.442493	0.3310	0.1927	0.4390
	N(1,0.5)	N(1,0.5)	0.373389	0.2481	0.1539	0.3923
	N(0,2)	N(1,1)	0.3406	0.2087	0.0725	0.2693
	N(0,2)	N(1,0.5)	0.364364	0.2372	0.1016	0.3188
	N(1,2)	N(1,0.5)	0.395651	0.2748	0.1103	0.3321
	N(1,2)	N(2,3)	0.341339	0.2096	0.0725	0.2693
	N(1,2)	N(2,2)	0.345068	0.2141	0.0926	0.3043
	N(1,2)	N(2,1)	0.354583	0.2255	0.1466	0.3829
	N(1,2)	N(2,0.5)	0.361739	0.2341	0.1748	0.4181
	N(1,1)	N(2,0.5)	0.579117	0.4949	0.2926	0.5409
	N(1,0.75)	N(2,0.5)	0.659266	0.5911	0.3798	0.6163
	N(2,1.5)	N(4,2)	0.659266	0.5911	0.3798	0.6163
	N(0,1)	N(3,1.25)	0.473685	0.3684	0.3397	0.5829
	N(1,1)	N(3,1.25)	0.660180	0.5922	0.3913	0.6255
	N(1,1.25)	N(3,1.25)	0.609918	0.5319	0.3354	0.5792
	N(1,1)	N(3,1)	0.684116	0.6209	0.43	0.6557
	N(1,1)	N(4,1)	0.743529	0.6922	0.5316	0.7291
	N(0,1)	N(0.1,0.5)	0.216916	0.0603	0.0106	0.1030
	N(2,1)	N(6,1)	0.919012	0.9028	0.8209	0.9060
	N(3,1)	N(9,1)	0.983042	0.9796	0.96	0.9798
	N(4,1)	N(16,1)	0.997661	0.9972	0.9944	0.9972
	N(0,1)	N(10,1)	0.5	0.4	0.4	0.6324
	N(0,10)	N(0,100)	0.249842	0.0998	0.0821	0.2865
	N(0,0.01)	N(0,0.001)	0.249984	0.1	0.0981	0.3132
	N(8,1)	N(16,1)	1	1	1	1

Table 7.3: Behavior of Distance Index Measures for $X_1 \sim Exp(\lambda_1)$, $X_2 \sim Exp(\lambda_2)$, $X_3 \sim Exp(\lambda_3)$

F_1	F_2	F_3	$p_{\max}$	γ_m	γ_{ss}	γ_{sd}
Exp(1)	Exp(1)	Exp(1)	0.166667	0	0	0
	Exp(1)	Exp(2)	0.25	0.1	0.033	0.1826
	Exp(1)	Exp(0.5)	0.266667	0.12	0.0373	0.1932
	Exp(0.25)	Exp(0.5)	0.380952	0.2571	0.0979	0.3129
	Exp(1)	Exp(4)	0.333333	0.2	0.112	0.3347
	Exp(3)	Exp(4)	0.375	0.25	0.1118	0.3344
	Exp(3)	Exp(6)	0.45	0.34	0.1582	0.3978
	Exp(3)	Exp(9)	0.519231	0.4231	0.2163	0.4651
	Exp(0.33)	Exp(0.66)	0.335008	0.202	0.0683	0.2613
	Exp(1)	Exp(0.66)	0.224271	0.0691	0.0127	0.1126
	Exp(1)	Exp(0.33)	0.322695	0.1872	0.0892	0.2986
	Exp(1)	Exp(0.25)	0.355556	0.2267	0.1298	0.3602
	Exp(1)	Exp(0.1)	0.4329	0.3195	0.2556	0.5056
	Exp(1)	Exp(0.01)	0.492587	0.3911	0.3824	0.6184
	Exp(2)	Exp(0.01)	0.657873	0.5894	0.4506	0.6712
	Exp(4)	Exp(0.01)	0.790498	0.7486	0.5975	0.7730
	Exp(10)	Exp(0.01)	0.899272	0.8791	0.7804	0.8834
	Exp(4)	Exp(0.25)	0.609524	0.5314	0.3142	0.5606
	Exp(4)	Exp(16)	0.609524	0.5314	0.3142	0.5606
	Exp(4)	Exp(8)	0.492308	0.3908	0.2041	0.4518
	Exp(6)	Exp(11)	0.457143	0.3486	0.2145	0.4632
	Exp(0.95)	Exp(1.05)	0.179487	0.0154	0.0006	0.0242
	Exp(0.75)	Exp(1.25)	0.238095	0.0857	0.0151	0.1228
	Exp(0.05)	Exp(25)	0.913992	0.8968	0.8067	0.8982
	Exp(1.5)	Exp(0.5)	0.333333	0.2	0.0673	0.2595

Table 7.4: Behavior of Distance Index Measures for $X_1 \sim Logis(0,1)$, $X_2 \sim Logis(d_1, s_1)$ and $X_3 \sim Logis(d_2, s_2)$

F_1	F_2	F_3	$\hat{p}_{\max}$	$\hat{\gamma}_m$	$\hat{\gamma}_{ss}$	$\hat{\gamma}_{sd}$
L(0,1)	L(0,1)	L(0,1)	0.166667	0.0038	0.0001	0.0094
	L(1,1)	L(2,1)	0.3768	0.2521	0.0946	0.3076
	L(0,2)	L(0,3)	0.2268	0.0722	0.0162	0.1273
	L(1,2)	L(2,3)	0.2684	0.1220	0.0379	0.1946
	L(2,2)	L(4,3)	0.3832	0.2598	0.0985	0.3138
	L(2,3)	L(4,3)	0.3028	0.1634	0.0729	0.2699
	L(2,1)	L(4,3)	0.4994	0.3992	0.1787	0.4227
	L(2,0.5)	L(4,3)	0.5618	0.4742	0.2442	0.4942
	L(0,1)	L(2,1)	0.3454	0.2145	0.1168	0.3417
	L(0,1)	L(2,2)	0.3121	0.1745	0.0730	0.2701
	L(0,1)	L(2,0.5)	0.3737	0.2485	0.1608	0.4010
	L(3,2)	L(6,3)	0.4995	0.3994	0.1930	0.4393
	L(3,2)	L(6,4)	0.4786	0.3743	0.1664	0.4080
	L(1,2)	L(2,1)	0.2812	0.1375	0.0753	0.2744
	L(1,2)	L(2,0.5)	0.3240	0.1888	0.1089	0.3300
	L(1,1)	L(2,0.5)	0.3927	0.2712	0.1251	0.3538
	L(1,0.75)	L(2,0.5)	0.4587	0.3505	0.1585	0.3982
	L(2,1.5)	L(4,2)	0.4583	0.35	0.1575	0.3969
	L(0,1)	L(3,1.25)	0.3974	0.2769	0.1872	0.4326
	L(1,1)	L(3,1.25)	0.4677	0.3613	0.1705	0.4129
	L(1,1.25)	L(3,1.25)	0.4248	0.3098	0.1457	0.3816
	L(1,1)	L(3,1)	0.4843	0.3812	0.1932	0.4395
	L(1,1)	L(4,1)	0.5635	0.4762	0.2945	0.5427
	L(2,1)	L(6,1)	0.7443	0.6932	0.5121	0.7156
	L(3,1)	L(9,1)	0.8801	0.8562	0.7433	0.8622
	L(4,1)	L(16,1)	0.9470	0.9364	0.8796	0.9378
	L(0,1)	L(10,1)	0.5071	0.4086	0.3992	0.6319
	L(0,10)	L(0,100)	0.2514	0.1017	0.0806	0.2838
	L(0,0.01)	L(0,0.001)	0.2577	0.1092	0.0980	0.3130
	L(8,1)	L(16,1)	0.9956	0.9947	0.9895	0.9947
	L(15,1)	L(30,1)	1	1	1	1

Table 7.5: Behavior of Distance Index Measures for $X_1 \sim GEV(0,1,\psi)$, $X_2 \sim GEV(d_1,s_1,\psi_1)$ and $X_3 \sim GEV(d_2,s_2,\psi_2)$

F_1	F_2	F_3	$\hat{p}_{\max}$	$\hat{\gamma}_m$	$\hat{\gamma}_{ss}$	$\hat{\gamma}_{sd}$
G(0,1,0)	G(1,1,0)	G(2,1,0)	0.4909	0.3891	0.1904	0.4364
	G(0,2,0)	G(0,3,0)	0.2317	0.0781	0.0203	0.1484
	G(1,2,0)	G(2,3,0)	0.3778	0.2533	0.0965	0.3107
	G(2,2,0)	G(4,3,0)	0.5665	0.4798	0.2746	0.5240
	G(2,1,0)	G(4,3,0)	0.6881	0.6258	0.4102	0.6404
	G(2,0.5,0)	G(4,3,0)	0.7348	0.6817	0.4752	0.6894
	G(2,0.5,0)	G(4,0.5,0)	0.8766	0.8519	0.7328	0.8560
	G(0,1,0)	G(2,1,0)	0.3990	0.2789	0.1961	0.4428
G(0,1,0.5)	G(1,1,0.5)	G(2,1,0.5)	0.4460	0.3352	0.1459	0.3820
	G(1,1.5,0)	G(2,2,-.5)	0.3580	0.2296	0.0696	0.2638
	G(1,2,0)	G(2,1.5,-.5)	0.2878	0.1453	0.0618	0.2487
	G(1,2,-.5)	G(2,1.5,0)	0.3502	0.2202	0.0841	0.2899
	G(1,1.5,-.5)	G(2,2,0)	0.4074	0.2889	0.1025	0.3202
G(0,1,0)	G(1,1.5,-.5)	G(2,2,.5)	0.4584	0.3500	0.1606	0.4008
	G(1,2,-.5)	G(2,1.5,.5)	0.4053	0.2864	0.1425	0.3775
	G(1,2,.5)	G(2,1.5,-.5)	0.3535	0.2242	0.1085	0.3294
	G(1,1.5,.5)	G(2,2,-.5)	0.3582	0.2298	0.1029	0.3207
G(0,1,-.5)	G(1,1,-.5)	G(2,1,-.5)	0.5784	0.4941	0.2813	0.5303
	G(2,1,-.5)	G(4,1,-.5)	0.8748	0.8497	0.7274	0.8529
	G(1,1.5,0)	G(2,2,0.5)	0.4774	0.3729	0.1940	0.4405
	G(1,1.5,.5)	G(2,2,0)	0.4179	0.3015	0.1600	0.4000
	G(1,2,0)	G(2,1.5,.5)	0.4084	0.2900	0.1803	0.4246
	G(1,2,.5)	G(2,1.5,0)	0.3734	0.2480	0.1652	0.4064
	G(0,1,0)	G(0,1,.5)	0.2148	0.0577	0.0089	0.0942
G(0,1,0)	G(0,1,-1)	G(0,1,1)	0.2534	0.1040	0.0213	0.1460
G(0,1,0)	G(0,1,-.2)	G(0,1,.2)	0.1869	0.0243	0.0016	0.0393
G(0,1,.2)	G(1,1,.2)	G(2,1,.2)	0.4692	0.3630	0.1683	0.4102
G(0,1,-.2)	G(1,1,-.2)	G(2,1,-.2)	0.5204	0.4245	0.2213	0.4704
G(0,1,-1)	G(1,1,-1)	G(2,1,-1)	0.6419	0.5703	0.3495	0.5912
G(0,1,1)	G(1,1,1)	G(2,1,1)	0.4197	0.3037	0.1229	0.3506
G(0,1,0)	G(0,10,0)	G(0,20,0)	1	1	1	1

Looking at the behavior of these index measures, a few summary observations are worth noting. For all the examples that we looked into, $\gamma_{sd} \geq \gamma_m \geq \gamma_{ss}$. Also, since, $\gamma_{ss} = \gamma_{sd}^2$, for most of these examples, γ_{ss} seems to be much lower than the other two index measures. This indicates that possibly γ_{sd} is a safer choice among these two related index measures. The maximum possible value of 1 for all these index measures is approached when there are large differences in the location parameters. However, the difference in location parameters at which these measures are very close to 1, depends on the parametric distributions involved.

In the future, I plan to investigate statistical inference for these index measures.

Chapter 8: A Family of Hypothesis Tests

Recall that in Chapter 3, motivated by the well known and important issue of testing for *Hypervolume under Manifold* (HUM) of a ROC Manifold, I proposed testing the following hypotheses

$$H_0 : F_1 = F_2 = ... = F_K$$

Vs.

$$H_A : \Pr(X_1 < X_2 < ... < X_K) > \frac{1}{K!}$$

(8.1)

where $\{X_i, i = 1,2,...,K\}$ are independently distributed with absolutely continuous distribution functions $\{F_i; i = 1,2,...,K\}$.

Much can be learned about the joint behavior of K distributions from their precedence probabilities, an approach that has not yet been looked at very much and (8.1) can be viewed as a particular representative of a family of tests described in this chapter. Simply put, I propose a new way to compare K distributions based on independent random samples using a new paradigm. A very old problem in the literature of statistics is what is known as the *K*-sample problem of testing for inequality among the distributions in the above setting.

Clearly all of the distributions are equal and continuous, each precedence probability $\{p(\xi_i); i = 1, 2, ..., K!\}$ equals $1/K!$. It is not known if the converse is true for $K \geq 3$. However, we have not yet found an example where any departure from equality of distributions does not change at least one of the $K!$ precedence probabilities if $K \geq 3$ and all the distributions are continuous.

On the other hand, for $K = 2$, and symmetric location-scale families of distributions, if the locations are the same, both precedence probabilities equal ½ even if the distributions differ in scale. If there is prior reason to believe that one or a relatively small group of orderings is dominant, rather than testing against the usual alternative that at least two of the distributions are not identical, it would make more sense to test for a more targeted alternative hypothesis. The *HUM* under the ROC manifold is an example of this type of directed test. Also, consider an experiment where X_i is a score corresponding to a treatment $T_i; i = 1, 2, 3$ and it is only suspected that X_2 and X_3 tend to be larger than X_1; whereas the ordering of X_2 and X_3 is unknown. In such a case it might be a good idea to just test whether at least one of the precedence probabilities $P(X_1 < X_2 < X_3)$ and $P(X_1 < X_3 < X_2)$ is greater than $1/K!$.

Generalizing this approach, I propose using alternative hypotheses for the K-sample problem based on precedence probabilities. Since, $\sum_{i=1}^{K!} p(\xi_i) = 1$, it is sufficient to test for at most $(K! - 1)$ precedence probabilities. I will henceforth refer to this formulation as testing for precedence.

Keeping the above examples in mind, I begin with a "precedence probability"-based test for the K-sample problem whose hypotheses are

$$H_0 : F_1 = F_2 = ... = F_K$$

Vs.

$$H_A : \{\text{at least one of } p(\xi_i) \neq 1/K!; i = 1,2,...,m\}$$

(8.2)

where $\{p(\xi_i); i = 1,2,...,m\}$ is an apriori selected set of m precedence probabilities with $1 \leq m \leq (K!-1)$.

Note that, when $m = 1$, (8.2) is a two-sided version of the alternative hypotheses in (8.1).

Also note that, when $m = K!-1$,

$$H_A : \{\text{at least one of } p(\xi_i) \neq 1/K!; i = 1,2,...,K!-1\}$$

$$\Rightarrow \tilde{H}_A : F_i \neq F_j, \text{ for at least one pair } (i, j)$$

This test, henceforth called *Precedence Omnibus Test (POT)*, is treated in Chapter 9 along with other POT-type tests for $m < K!-1$. Please note that, inequality in two precedence probabilities implies inequality of at least a pair of

distributions. One-sided tests with ordered alternatives are developed in Chapter 11.

Chapter 9: Rank Tests for Precedence Probabilities

In this chapter, I will propose a new test based on U-statistics, which were discussed in some detail in Chapter 2, for the hypotheses given in (8.2). Asymptotic normality of the new test statistic will follow from known results on the limiting distribution of U-statistics and the test will be based on a UMVUE (Uniformly Minimum Variance Unbiased Estimator) for a family of distributions containing all continuous distributions.

The U-statistic approach provides a very intuitive way of estimating precedence probabilities that leads to tests of hypotheses. Again, consider a set of K independent random variables $(X_1, X_2, ..., X_K)$ with continuous probability density functions (p.d.f.) denoted by $(f_1, f_2, ..., f_K)$ and distribution functions $(F_1, F_2, ..., F_K)$. Let, $\{X_{ij}; j = 1, 2, ..., n_i\}$ be n_i independent observations from the underlying distribution function $F_i; i = 1, 2, ..., K$. Then, let us consider the pooled N dimensional data vector $\mathbf{X} = \{X_{ij}; j = 1, 2, ..., n_i; i = 1, 2, ..., K\}$, where $N = \sum_{i=1}^{K} n_i$.

Recall that $\{\xi_i = (\zeta_{i1}, \zeta_{i2}, ..., \zeta_{iK}); i = 1, 2, ..., K!\}$ denote the set of permutations of the first K integers $(1, 2, ..., K)$. To simplify the notation, without any loss of generality, first consider estimating $p(\xi_1) = P(X_1 < X_2 < ... < X_K)$. The first step is

to provide a good estimator for this parameter. The following U-statistic of degree $(1,1,\ldots,1)$ is an unbiased estimator of $p(\xi_1)$,

$$U_1 = \frac{1}{n_1 n_2 \ldots n_K} \sum_{i_1=1}^{n_1} \sum_{i_2=1}^{n_2} \ldots \sum_{i_K=1}^{n_K} I(X_{1i_1} < X_{2i_2} < \ldots < X_{Ki_K}) . \tag{9.1}$$

Unbiased estimators $\{U_i; i = 1,2,\ldots,K!\}$ of the other precedence probabilities are defined in a similar manner. As described earlier, interest may lie in inference for $\{p(\xi_i); i = 1,2,\ldots,m\}$, an "a priori" selected set of m precedence probabilities with $1 \le m \le (K!-1)$. Let $\underset{\sim}{U}_m = (U_1, U_2, \ldots, U_m)$ be the corresponding vector of U-statistic estimators, with

$$U_i = \frac{1}{n_1 n_2 \ldots n_K} \sum_{j_1=1}^{n_{\zeta_{j1}}} \sum_{j_2=1}^{n_{\zeta_{j2}}} \ldots \sum_{j_K=1}^{n_{\zeta_{jK}}} I(X_{\zeta_{j1}j_1} < X_{\zeta_{j2}j_2} < \ldots < X_{\zeta_{jK}j_K}) . \tag{9.2}$$

Let V_m be the variance-covariance matrix of $\underset{\sim}{U}_m$ so that $v_m(i,j)$, the (i,j)-th element of V_m is $Cov(U_i, U_j)$, the covariance between U_i and U_j; 1_m be a column vector of dimension m and with all its elements being 1. Note that the dependence of V_m on the sample sizes $\{n_i, i = 1,2,\ldots,m\}$ is suppressed for the sake of simplicity.

Using the properties of U–statistics described in Chapter 2,

$$E_{H_0}(\underset{\sim}{U}_m) = \left(\frac{1}{K!}\right) 1_m . \tag{9.3}$$

and asymptotically, $\left(\underset{\sim}{U}_m - \left(\frac{1}{K!}\right)1_m\right) \sim N(0_m, V_m)$, where 0_m is a vector of length m with all its elements being 0, provided $\frac{n_i}{N} \to \nu_i; 0 < \nu_i < 1; \sum_{i=1}^{K} \nu_i = 1$ as $N \to \infty$.

Motivated by the asymptotic normality of $\underset{\sim}{U}_m = (U_1, U_2, ..., U_m)$ under H_0, I propose the following statistic for testing (8.2):

$$\chi^2_{m(K)} = \left(\underset{\sim}{U}_m - \left(\frac{1}{K!}\right)1_m\right)' V_m^{-1} \left(\underset{\sim}{U}_m - \left(\frac{1}{K!}\right)1_m\right), \tag{9.4}$$

The asymptotic normality described above implies that the asymptotic distribution of our proposed test statistic under H_0 in (8.2) is a chi-square distribution with m degrees of freedom. An asymptotic size-α test is therefore obtained by the following rejection rule. Reject H_0 given in (8.2) if and only if $\chi^2_{m(K)} > \mathrm{X}^2_{m,\alpha}$ where $\mathrm{X}^2_{m,\alpha}$ is such that $P(X > \mathrm{X}^2_{m,\alpha}) = \alpha$, and where the random variable X has a chi-square distribution with m degrees of freedom. Properties of this class of tests will be discussed below.

Stepping back a bit, since the hypotheses H_0 and H_A in (8.2) are invariant with respect to strictly increasing, continuous transformations, we are led to consider tests based on the rank vectors $\mathbf{R} = \{\underline{R}_i;\ i = 1, 2, ..., K\}$, where $\underline{R}_i$ are the ranks of $\{X_{ij}, j = 1, 2, ..., n_i\}$ in the pooled N dimensional data vector $\underline{Y} = \{X_{ij}\}$. The proposed U-statistic estimator U_i is unbiased for $p(\xi_i); i = 1, 2, ..., K!$ and $U_i(\underline{Y}) = U_i(\mathbf{R})$. Clearly, the test statistic introduced in (9.3) is rank-based, but not a

linear rank statistic. Consequently, having observed data $\mathbf{X} = \mathbf{x} \equiv (x_{ij})$, a naturally appealing class of permutation test $\{\psi_c\}$ of (8.2) is given by

$$\psi_c(\underline{y}) = 1 \text{ if } \chi^2_{m(K)}(\underline{y}) \geq c,$$

$$= 0 \text{ otherwise,} \qquad (9.5)$$

where $c \geq 1/K!$ is a constant. Let $\prod = \{\pi_V\}$ denote the $N!$ permutations of the components of $\mathbf{x}$. The size of such a test is given by

$$E(\psi_c(\mathbf{X}) \mid H_0) = \#\{v;\ \chi^2_{m(K)}(\pi_v(\mathbf{x})) \geq c\} / N!.$$

Following an argument given in Bhapkar (1961), I plan to show in the future that this test is consistent (the power goes to 1 if null hypothesis is false) against all alternatives in H_a.

Choosing the constant c so that $\psi_c(\mathbf{x})$ is close to some desired value can be numerically difficult. Instead, a permutation *p-value* is given by

$$\textit{p-value} = \#\{v;\ \chi^2_{m(K)}(\pi_v(\underline{y})) \geq \chi^2_{m(K)}(\underline{y})\} / N!.$$

Then, an approximate size-α test may be carried out by rejecting the null hypothesis only if $\textit{p-value} \leq \alpha$. To further reduce the amount of computation, instead of computing $\chi^2_{m(K)}(\pi_v(\underline{y}))$ for all permutations $\prod = \{\pi_V\}$, following Efron and Tibshirani (1993) and Davison and Hinkley (1997) a permutation *p-value* may be approximated by sampling from $\prod = \{\pi_V\}$ as follows.

(i) For a large, positive integer L, independently and randomly select L permutations $\prod^* = \{\pi_v^*, v = 1,2,...,L\}$.

(i) Compute $U_K(\pi_v^*(\mathbf{x}))$, for each permutation π_v^* in Π^*

(ii) Estimate a *p-value* by then calculating the empirical p-value as

$$\hat{p} = \frac{1 + \#\{\chi^2_{m(K)}(\pi_v^*(\underline{y})) > \chi^2_{m(K)}(\underline{y})\}}{L+1} \quad . \tag{9.6}$$

Rejecting H_0 if $\hat{p}$ at is at most α yields an approximate size-α test.

The first step in being able to use (9.3) is to derive expressions for calculating the elements of V_m. To that end, let $p(\xi_i) = P(X_{\alpha_1} < X_{\alpha_2} < ... < X_{\alpha_K})$ and $p(\xi_j) = P(X_{\beta_1} < X_{\beta_2} < ... < X_{\beta_K})$ where $\zeta_{il} = \alpha_l$, $\zeta_{jl} = \beta_l; \forall l = 1,2,...,K$ and both $\xi_i \in \xi, \xi_j \in \xi$; ξ being the collection of all permutations of $\{1,2,...,K\}$. Note again that U_i and U_j are unbiased *U*-statistic estimates for $p(\xi_i)$ and $p(\xi_j)$, respectively. When the null hypothesis in (8.2) is true, using (9.4),

$$Cov(U_i, U_j) = E(U_i U_j) - (1/K!)^2 . \tag{9.7}$$

So, to obtain the covariance, we need to calculate $E(U_i U_j)$ when the null hypothesis is true, denoted by $E_{H_0}(U_i U_j)$.

By definition,

$$E_{H_0}(U_i U_j) = E\left(\frac{1}{n_0^2} \sum_{i_1=1}^{n_{\alpha_1}} \ldots \sum_{i_K=1}^{n_{\alpha_K}} \sum_{j_1=1}^{n_{\beta_1}} \ldots \sum_{j_K=1}^{n_{\beta_K}} I(X_{\alpha_1 n_{i_1}} < \ldots < X_{\alpha_K n_{i_K}}) I(X_{\beta_1 n_{j_1}} < \ldots < X_{\beta_K n_{j_K}}) \right); \quad (9.8)$$

with $n_0 = \prod_{i=1}^{K} n_i$.

Now, let

$$c_{\alpha(l),\beta(r)} = I(X_{\alpha_1 l_1} < \ldots < X_{\alpha_K l_K}) I(X_{\beta_1 r_1} < \ldots < X_{\beta_K r_K}) \quad (9.9)$$

and $c_{\alpha(l),\beta(r),s,b,b_*} = I(X_{\alpha_1 l_1} < \ldots < X_{\alpha_K l_K}) I(X_{\beta_1 r_1} < \ldots < X_{\beta_K r_K})$ with $\alpha_b = \beta_{b_*}$ and $l_b = r_{b_*}$ for exactly s positions at $b = (d_1, \ldots, d_s)$ and $b_* = (g_1, \ldots, g_s)$. The indicator function $c_{\alpha(l),\beta(r),s,b,b_*}$ has the value 1 if and only if $c_{\alpha(l)} = I(X_{\alpha_1 l_1} < \ldots < X_{\alpha_K l_K})$ and $c_{\beta(r)} = I(X_{\beta_1 r_1} < \ldots < X_{\beta_K r_K})$ have s common observations occurring from distributions $\{F_j; j = \alpha_{d_1}, \ldots, \alpha_{d_s}\}$. Note that, in this scenario, $\alpha_{d_i} = \beta_{g_i}; i = 1, \ldots, s$. However, d_i is not necessarily the same as g_i.

Also, let us define $c_{\alpha(l),\beta(r),0} = I(X_{\alpha_1 l_1} < \ldots < X_{\alpha_K l_K}) I(X_{\beta_1 r_1} < \ldots < X_{\beta_K r_K})$ such that $l_i \neq r_i; i = 1, 2, \ldots, K$. Specifically, there is no common observation in $c_{\alpha(l)} = I(X_{\alpha_1 l_1} < \ldots < X_{\alpha_K l_K})$ and $c_{\beta(r)} = I(X_{\beta_1 r_1} < \ldots < X_{\beta_K r_K})$ coming out of any of the underlying distributions. To proceed, we define the *concordance* between vectors as follows.

Definition: Vectors $g = (g_1, \ldots, g_s)$ and $h = (h_1, \ldots, h_s)$ are concordant if and only if $(h_1 < \ldots < h_s) \Leftrightarrow (g_1 < \ldots < g_s)$.

Then, we have the following results presented in a theorem.

Theorem 9.1: Consider, $c_{\alpha(l),\beta(r)} = I(X_{\alpha_1 l_1} < ... < X_{\alpha_K l_K}) I(X_{\beta_1 r_1} < ... < X_{\beta_K r_K})$ with $\alpha_b = \beta_{b_*}$ and $l_b = r_{b_*}$ for exactly s positions at $b = (d_1, ..., d_s)$ and $b_* = (g_1, ..., g_s)$. Then, if the null hypothesis mentioned in (8.2) is correct,

(i) provided b and b_* are concordant and $(g_1 < ... < g_s)$, for $s = 1, 2, ..., K$;

$$E_{H_0}(c_{\alpha(l),\beta(r),s,b,b_*}) = C_{d_1 - 1}^{g_1 + d_1 - 2} \left\{ \prod_{i=1}^{s-1} C_{d_{i+1} - d_i - 1}^{g_{i+1} + d_{i+1} - d_i - g_i - 2} \right\} C_{K - d_s}^{2K - g_s - d_s} \Big/ (2K - s)! \; ; \tag{9.10}$$

(ii) If b and b_* are not concordant, for $s = 1, 2, ..., K$;

$$E_{H_0}(c_{\alpha(l),\beta(r),s,b,b_*}) = 0 \; ; \tag{9.11}$$

(iii) For $s = 0$;

$$E_{H_0}(c_{\alpha(l),\beta(r),0}) = \left(\frac{1}{K!} \right)^2 \; . \tag{9.12}$$

Proof: (i) Clearly, there are $(2K - s)$ unique observations involved in $c_{\alpha(l),\beta(r)}$., which can be arranged in $(2K - s)!$ ways and, under the null hypothesis, all of these arrangements are equally likely. Now, note that $X_{\alpha_{d_1} l_{d_1}} = X_{\beta_{g_1} r_{g_1}}$.

Then, $\{(X_{\alpha_1 l_1},...,X_{\alpha_{d_1-1} l_{d_1-1}});(X_{\beta_1 r_1},...,X_{\beta_{g_1-1} r_{g_1-1}})\}$ can be arranged in $C_{d_1-1}^{g_1+d_1-2}$ ways so that $(X_{\alpha_1 l_1} < ... < X_{\alpha_{d_1-1} l_{d_1-1}})$ and $(X_{\beta_1 r_1} < ... < X_{\beta_{g_1-1} r_{g_1-1}})$ and under the null hypothesis in (8.2), all of these are equally likely.

In a similar manner, $\{(X_{\alpha_{d_1+1} l_{d_1+1}},...,X_{\alpha_{d_2-1} l_{d_2-1}});(X_{\beta_{g_1+1} r_{g_1+1}},...,X_{\beta_{g_2-1} r_{g_2-1}})\}$ can be arranged in $C_{d_2-1-d_1}^{g_2-1-g_1+d_2-1-d_1}$ ways with $(X_{\alpha_{d_1+1} l_{d_1+1}} < ... < X_{\alpha_{d_2-1} l_{d_2-1}})$ and $(X_{\beta_{g_1+1} r_{g_1+1}} < ... < X_{\beta_{g_2-1} r_{g_2-1}})$. If the null hypothesis is true, these arrangements are equally likely.

It follows in a similar fashion that

$$E_{H_0}(c_{\alpha(l),\beta(r),s,b,b_*}) = C_{d_1-1}^{g_1+d_1-2}\left\{\prod_{i=1}^{s-1} C_{d_{i+1}-d_i-1}^{g_{i+1}+d_{i+1}-d_i-g_i-2}\right\} C_{K-d_s}^{2K-g_s-d_s} \Big/ (2K-s)!.$$

Provided b and b_* are concordant and $(g_1 < ... < g_s)$.

(ii) Without any loss of generality, consider $d_1 = g_2$ and $d_2 = g_1$.

Then, $X_{\alpha_{d_1} l_{d_1}} = X_{\beta_{g_2} r_{g_2}}$ and $X_{\alpha_{d_2} l_{d_2}} = X_{\beta_{g_1} r_{g_1}}$. Since, $d_1 < d_2$ and $g_1 < g_2$, it is impossible to obtain observations in any way so that all the conditions are satisfied (i.e.; $X_{\alpha_{d_1} l_{d_1}}$ cannot be both greater and smaller than $X_{\alpha_{d_2} l_{d_2}}$) which leads to the result given in equation (9.11).

(iii) Since, $c_{\alpha(l),\beta(r)}$ has no common observation present in both parts $I(X_{\alpha_1 l_1} < \ldots < X_{\alpha_K l_K})$ and $I(X_{\beta_1 r_1} < \ldots < X_{\beta_K r_K})$; with the assumption that all observations are independent; we have the following result.

$$E_{H_0}(c_{\alpha(l),\beta(r)}) = E_{H_0}(I(X_{\alpha_1 l_1} < \ldots < X_{\alpha_K l_K}))E_{H_0}(I(X_{\beta_1 r_1} < \ldots < X_{\beta_K r_K}))$$

$$= (1/K!)(1/K!) \qquad \text{[Using (9.4)]}$$

$$= (1/K!)^2$$

This completes the result mentioned in (9.12) and leads to the next result.

Theorem 9.2: If the null hypothesis in (8.2) is true;

$$E_{H_0}(U_i U_j) = \left(\frac{\prod_{i=1}^{K}(n_i - 1)}{n_0} \right) \left(\frac{1}{K!} \right)^2 + \sum_{s=1}^{K} \sum_{b,b* \in B_S} \left(\frac{\prod_{\alpha_i \notin b}(n_{\alpha_i} - 1)}{n_0} \right) \left(\frac{C_{d_1 - 1}^{g_1 + d_1 - 2} \left\{ \prod_{i=1}^{s-1} C_{d_{i+1} - d_i - 1}^{g_{i+1} + d_{i+1} - d_i - g_i - 2} \right\} C_{K - d_s}^{2K - g_s - d_s}}{(2K - s)!} \right) \tag{9.13}$$

where B_s is the collection of all concordant vectors of length s and $n_0 = \prod_{i=1}^{K} n_i$.

Proof: Consider,

$$E_{H_0}(U_iU_j)=\left(\sum_{l\neq r}E_{H_0}(c_{\alpha(l),\beta(r),0})+\sum_{s=1}^{K}\sum_{b,b_*\in B_S}\sum_{l,r}E_{H_0}(c_{\alpha(l),\beta(r),s,b,b_*})+\sum_{s=1}^{K}\sum_{b,b_*\notin B_S}\sum_{l,r}E_{H_0}(c_{\alpha(l),\beta(r),s,b,b_*})\right)\Big/ n_0^2 \tag{9.14}$$

Using (9.11),

$$\sum_{s=1}^{K}\sum_{b,b_*\notin B_S}\sum_{l,r}E_{H_0}(c_{\alpha(l),\beta(r),s,b,b_*})=0\,. \tag{9.15}$$

Consider the case where there are no common observations in $c_{\alpha(l)}=I(X_{\alpha_1 l_1}<\ldots<X_{\alpha_K l_K})$ and $c_{\beta(r)}=I(X_{\beta_1 r_1}<\ldots<X_{\beta_K r_K})$. There are $n_i(n_i-1)$ ways to select two different observations from distribution $F_i; i=1,\ldots,K$. So, there are $\prod_{i=1}^{K}n_i(n_i-1)$ ways of selecting two observations from each of K distributions without obtaining any common observation.

Hence, $\sum_{l\neq r}E_{H_0}(c_{\alpha(l),\beta(r),0})=\left(\prod_{i=1}^{K}n_i(n_i-1)\right)E_{H_0}(c_{\alpha(l),\beta(r),0})$

$$=\left(\prod_{i=1}^{K}n_i(n_i-1)\right)\left(\frac{1}{K!}\right)^2 \qquad \text{[Using (9.12)]} \tag{9.16}$$

Now, let us consider the case of $c_{\alpha(l)}=I(X_{\alpha_1 l_1}<\ldots<X_{\alpha_K l_K})$ and $c_{\beta(r)}=I(X_{\beta_1 r_1}<\ldots<X_{\beta_K r_K})$ having s common observations at positions $b=(d_1,\ldots,d_s)$ and $b_*=(g_1,\ldots,g_s)$, respectively, from distributions $\{F_j; j=\alpha_{d_1},\ldots,\alpha_{d_s}\}$.

The number of ways to obtain 2 observations from each of K distributions such that the observations are same from distributions $\{F_j; j=\alpha_{d_1},...,\alpha_{d_s}\}$ but not same for any other distributions is $\left(\prod_{i=1}^{K} n_i \prod_{\alpha_i \notin b}(n_{\alpha_i}-1)\right)$.

So, $\sum_{s=1}^{K}\sum_{b,b_* \in B_S}\sum_{l,r} E_{H_0}(c_{\alpha(l),\beta(r),s,b,b_*}) = \sum_{s=1}^{K}\sum_{b,b_* \in B_S}\left(n_0 \prod_{\alpha_i \notin b}(n_{\alpha_i}-1)\right) E_{H_0}(c_{\alpha(l),\beta(r),s,b,b_*})$

$$= \sum_{s=1}^{K}\sum_{b,b_* \in B_S}\left(n_0 \prod_{\alpha_i \notin b}(n_{\alpha_i}-1)\right)\left(\frac{C_{d_1-1}^{g_1+d_1-2}\left\{\prod_{i=1}^{s-1} C_{d_{i+1}-d_i-1}^{g_{i+1}+d_{i+1}-d_i-g_i-2}\right\} C_{K-d_s}^{2K-g_s-d_s}}{(2K-s)!}\right)$$

[Using (9.10)] (9.17)

Combining (9.14), (9.15), (9.16) and (9.17), we obtain the result in (9.13).

Now, we use the above results to illustrate how to obtain expressions for the elements of the variance-covariance matrix V.

Example 9.1: For $K=2$, consider, $p_2(\xi_1)=P(X_1<X_2)$ and $p_2(\xi_2)=P(X_2<X_1)$. Let, $U_{i(2)}$ be the U-statistic estimator for $p_2(\xi_i); i=1,2$ and consider the singular variance-covariance matrix of $U_{(2)}=(U_{1(2)},U_{2(2)})'$, denoted by V_2;

with $V_2(1,1)=Var(U_{1(2)})$; $V_2(2,2)=Var(U_{2(2)})$ and $V_2(1,2)=V_2(2,1)=Cov(U_{1(2)},U_{2(2)})$.

As mentioned earlier, for $K=2$ we only need to compute one of the two precedence probabilities since the precedence probabilities must add up to 1. Accordingly, consider $p_2(\xi_1)=P(X_1<X_2)$. We then obtain

$$V_2(1,1)=\frac{1}{4}\frac{1}{n_1n_2}+\frac{1}{12}\frac{(n_1+n_2-2)}{n_1n_2}=\frac{1}{12}\frac{(n_1+n_2+1)}{n_1n_2}, \qquad (9.18)$$

a well known result.

Example 9.2: For $K=3$ define the following precedence probabilities,

$p_3(\xi_1)=P(X_1<X_2<X_3)$; $p_3(\xi_2)=P(X_1<X_3<X_2)$; $p_3(\xi_3)=P(X_2<X_1<X_3)$;

$p_3(\xi_4)=P(X_2<X_3<X_1)$; $p_3(\xi_5)=P(X_3<X_1<X_2)$ and $p_3(\xi_6)=P(X_3<X_2<X_1)$.

In this setting we might be interested to either test for certain *a priori* specified precedence probabilities or to carry out a POT-type test for testing the equality of distributions.

Let, $U_{i(3)}$ be the U-statistic estimator for $p_3(\xi_i); i=1,2,...,6$. To carry out a POT-type test in this scenario, it suffices to consider a test statistic based on $U_{(3)}=(U_{1(3)},U_{2(3)},U_{3(3)},U_{4(3)},U_{5(3)})'$. Then, consider, the variance-covariance matrix V_3 with its (i,j)-th element being $V_3(i,j)=Cov(U_{i(3)},U_{j(3)}); i=1,2,...,5; j=1,2,...,5$.

We have the following results;

$$V_3(1,1)=\frac{1}{n_0}\left(\frac{1}{6}+\left(\frac{n_1+n_2+n_3-3}{12}\right)+\left(\frac{(n_2-1)(n_1+n_3-2)}{20}\right)+\left(\frac{(n_1-1)(n_3-1)}{30}\right)\right)-\frac{n_{-1}^*}{36};$$

$$V_3(1,2)=V_3(2,1)=\frac{1}{n_0}\left(\left(\frac{n_2+n_3-2}{24}\right)+\left(\frac{(n_2-1)(n_3-1)}{20}\right)+\left(\frac{(n_1-1)(n_2+n_3-2)}{40}\right)\right)-\frac{n_{-1}^*}{36};$$

$$V_3(1,3)=V_3(3,1)=\frac{1}{n_0}\left(\left(\frac{n_2+n_1-2}{24}\right)+\left(\frac{(n_2-1)(n_1-1)}{20}\right)+\left(\frac{(n_3-1)(n_2+n_1-2)}{40}\right)\right)-\frac{n_{-1}^*}{36};$$

$$V_3(1,4)=V_3(4,1)=\frac{1}{n_0}\left(\left(\frac{n_1-1}{24}\right)+\left(\frac{(n_3-1)(n_1-1)}{40}\right)+\left(\frac{(n_2-1)(n_3+n_1-2)}{120}\right)\right)-\frac{n_{-1}^*}{36};$$

$$V_3(1,5)=V_3(5,1)=\frac{1}{n_0}\left(\left(\frac{n_3-1}{24}\right)+\left(\frac{(n_3-1)(n_1+n_2-2)}{40}\right)+\left(\frac{(n_2-1)(n_1-1)}{120}\right)\right)-\frac{n_{-1}^*}{36};$$

$$V_3(2,2)=\frac{1}{n_0}\left(\frac{1}{6}+\left(\frac{n_1+n_2+n_3-3}{12}\right)+\left(\frac{(n_3-1)(n_1+n_2-2)}{20}\right)+\left(\frac{(n_1-1)(n_2-1)}{30}\right)\right)-\frac{n_{-1}^*}{36};$$

$$V_3(2,3)=V_3(3,2)=\frac{1}{n_0}\left(\left(\frac{n_2-1}{24}\right)+\left(\frac{(n_2-1)(n_1+n_3-2)}{40}\right)+\left(\frac{(n_3-1)(n_1-1)}{120}\right)\right)-\frac{n_{-1}^*}{36};$$

$$V_3(2,4)=V_3(4,2)=\frac{1}{n_0}\left(\left(\frac{(n_2-1)(n_1-1)}{30}\right)+\left(\frac{(n_3-1)(n_1+n_2-2)}{120}\right)\right)-\frac{n_{-1}^*}{36};$$

$$V_3(2,5)=V_3(5,2)=\frac{1}{n_0}\left(\left(\frac{(n_1+n_3-2)}{24}\right)+\left(\frac{(n_3-1)(n_1-1)}{20}\right)+\left(\frac{(n_2-1)(n_1+n_3-2)}{40}\right)\right)-\frac{n_{-1}^*}{36};$$

$$V_3(3,3)=\frac{1}{n_0}\left(\frac{1}{6}+\left(\frac{n_1+n_2+n_3-3}{12}\right)+\left(\frac{(n_1-1)(n_3+n_2-2)}{20}\right)+\left(\frac{(n_3-1)(n_2-1)}{30}\right)\right)-\frac{n_{-1}^*}{36};$$

$$V_3(3,4)=V_3(4,3)=\frac{1}{n_0}\left(\left(\frac{(n_1+n_3-2)}{24}\right)+\left(\frac{(n_3-1)(n_1-1)}{20}\right)+\left(\frac{(n_2-1)(n_1+n_3-2)}{40}\right)\right)-\frac{n_{-1}^*}{36};$$

$$V_3(3,5) = V_3(5,3) = \frac{1}{n_0}\left(\left(\frac{(n_3-1)(n_2-1)}{30}\right)+\left(\frac{(n_1-1)(n_2+n_3-2)}{120}\right)\right)-\frac{n_{-1}^*}{36};$$

$$V_3(4,4) = \frac{1}{n_0}\left(\frac{1}{6}+\left(\frac{n_1+n_2+n_3-3}{12}\right)+\left(\frac{(n_3-1)(n_1+n_2-2)}{20}\right)+\left(\frac{(n_1-1)(n_2-1)}{30}\right)\right)-\frac{n_{-1}^*}{36};$$

$$V_3(4,5) = V_3(5,4) = \frac{1}{n_0}\left(\left(\frac{(n_2-1)}{24}\right)+\left(\frac{(n_2-1)(n_1+n_3-2)}{40}\right)+\left(\frac{(n_1-1)(n_3-1)}{120}\right)\right)-\frac{n_{-1}^*}{36};$$

$$V_3(5,5) = \frac{1}{n_0}\left(\frac{1}{6}+\left(\frac{n_1+n_2+n_3-3}{12}\right)+\left(\frac{(n_1-1)(n_3+n_2-2)}{20}\right)+\left(\frac{(n_3-1)(n_2-1)}{30}\right)\right)-\frac{n_{-1}^*}{36}.$$

(9.19)

Note, $n_{-1}^* = \left(n_0-(n_1-1)(n_2-1)(n_3-1)\right)/n_0$ and $n_0 = n_1n_2n_3$.

This leads us to the POT described below. However, if we want to test for certain precedence probabilities or a linear combination of precedence probabilities, we can use these basic results described above as follows.

For example, if we want a test for a parameter, $p_3(\xi_1)-p_3(\xi_2)$, the test statistic could be based on $U_{1-2(3)} = U_{1(3)}-U_{2(3)}$. The variance of this test statistic is given by

$$V_{1-2(3)} = \frac{1}{n_1n_2n_3}\left(\frac{1}{3}+\left(\frac{7n_1+3n_2+3n_3-12}{36}\right)+\left(\frac{(n_1-1)(n_2+n_3-2)}{30}\right)\right)$$

(9.20)

where $N = n_1+n_2+n_3$; which can be used to calculate the test statistic.

Computing the covariance matrix for $K > 3$ is carried out in a similar manner.

Example 9.3: For $K = 4$, define the following precedence probabilities,

$p_4(\xi_1) = P(X_1 < X_2 < X_3 < X_4)$ and $p_4(\xi_2) = P(X_1 < X_2 < X_4 < X_3)$

and let $U_{i(4)}$ denote the U-statistic estimator for $p_4(\xi_i); i = 1,2$. Then, using the results above, we obtain

$$Var(U_{1(4)}) = \frac{1}{n_1 n_2 n_3 n_4}\left(\frac{1}{24} + \left(\frac{vs_1}{60}\right) + \left(\frac{vs_2}{120}\right) + \left(\frac{vs_3}{180}\right) + \left(\frac{vs_4}{252}\right) + \left(\frac{vs_5}{420}\right)\right) - \left(\frac{1}{24}\right)^2 n^*_{-1(4)},$$

with

$vs_1 = n_1 + n_2 + n_3 + n_4 - 4$,

$vs_2 = (n_3 - 1)(n_4 - 1) + (n_3 - 1)(n_2 - 1) + (n_2 - 1)(n_1 - 1)$,

$vs_3 = (n_2 - 1)(n_4 - 1) + (n_1 - 1)(n_4 - 1) + (n_3 - 1)(n_1 - 1)$,

$vs_4 = (n_2 - 1)(n_3 - 1)(n_1 + n_4 - 2)$,

$vs_5 = (n_1 - 1)(n_4 - 1)(n_2 + n_3 - 2)$,

$$n^*_{-1(4)} = \left(n_1 n_2 n_3 n_4 - (n_1 - 1)(n_2 - 1)(n_3 - 1)(n_4 - 1)\right) / n_1 n_2 n_3 n_4 \qquad (9.21)$$

and

$$Cov(U_{1(4)}, U_{2(4)}) = \frac{1}{n_1 n_2 n_3 n_4}\left(\left(\frac{cs_1}{120}\right)+\left(\frac{cs_2}{240}\right)+\left(\frac{cs_3}{360}\right)+\left(\frac{cs_4}{252}\right)+\left(\frac{cs_5}{420}\right)+\left(\frac{cs_6}{420}\right)\right)-\left(\frac{1}{24}\right)^2 n^*_{-1(4)},$$

with

$$cs_1 = (n_3 + n_4 - 2) + (n_3 - 1)(n_4 - 1),$$

$$cs_2 = (n_2 - 1)(n_4 - 1) + (n_3 - 1)(n_2 - 1),$$

$$cs_3 = (n_1 - 1)(n_4 - 1) + (n_3 - 1)(n_1 - 1),$$

$$cs_4 = (n_2 - 1)(n_4 - 1)(n_3 - 1),$$

$$cs_5 = (n_1 - 1)(n_4 - 1)(n_3 - 1),$$

$$cs_6 = (n_1 - 1)(n_2 - 1)(n_3 - 1) + (n_1 - 1)(n_2 - 1)(n_4 - 1). \quad (9.22)$$

These expressions can be used to construct tests for specified precedence probabilities. All the results in the above examples are calculated on the basis of theorem 9.2.

My next aim is to look at the asymptotic properties of the test in (9.3) carried out using its asymptotic chi square distribution. In particular, I am most interested in consistency. For convenient reference, I again give the test statistic in (9.4).

$$\chi^2_{m(K)} = \left(\underset{\sim}{U}_m - \left(\frac{1}{K!}\right)1_m\right)' V_m^{-1}\left(\underset{\sim}{U}_m - \left(\frac{1}{K!}\right)1_m\right). \quad (9.23)$$

Hoeffding (1948) shows that under any particular departure from equality of the *K*- distributions, U-statistic have a joint, asymptotic normal distribution provided that $\frac{n_i}{N} \to \nu_i; 0 < \nu_i < 1; \sum_{i=1}^{K} \nu_i = 1$ as $N \to \infty$. Specifically, for any particular distributions $\{F_i^{\theta}\}$ with $\underset{\sim}{\theta} = \{p(\xi_i); i = 1,2,...,m\}$, the vector of *U*-statistics that we base our tests on, referred to here as $\underset{\sim}{U}_{n_0}$, asymptotically follows a joint normal distribution such that approximately, for large samples,

$$\underset{\sim}{U}_{n_0} \sim N\left(\underset{\sim}{\theta}, \Sigma_{\underset{\sim}{\theta}}^{(\{n_i\})}\right), \tag{9.24}$$

However, the asymptotic behavior of *U*-statistics under contiguous alternatives in the setting of this chapter has not been fully worked out. It is possible, however, to speculate that they behave 'reasonably' in the following sense. Specifically, assume that under $\{F_i^{\theta}\}$, in probability, for any $\underset{\sim}{\theta}$,

$$N\Sigma_{\underset{\sim}{\theta}}^{(\{n_i\})} \to \Sigma_{\underset{\sim}{\theta}}\ , \tag{9.25}$$

$\Sigma_{\underset{\sim}{\theta}}$ being a positive definite matrix, which leads to the following result on consistency of the test in (9.4) .

Theorem 9.3: Assuming (9.24) and (9.25) hold and that the vector of precedence probabilities $\underset{\sim}{\theta} = \{p(\xi_i); i = 1,2,...,m\} \neq \underset{\sim}{1}/K!$, the power of the test in (9.4) goes to 1 if $\frac{n_i}{N} \to \nu_i; 0 < \nu_i < 1; \sum_{i=1}^{K} \nu_i = 1$.

Proof: Letting $\underset{\sim}{\theta}_0 = \underset{\sim}{1}/K!$, $\|.\|$ denote Euclidean distance and $\omega_{\max}$ denote the maximum eigen value of Σ_{θ_0}, from (9.24) and the law of large numbers, we have in $\{F_i^{\theta}\}$ probability that,

$$\begin{aligned}
\chi_m^2 &= \left(\underset{\sim}{U}_m - \underset{\sim}{\theta}_0\right)' \left(\Sigma_{\underset{\sim}{\theta}_0}^{(\{n_i\})}\right)^{-1} \left(\underset{\sim}{U}_m - \underset{\sim}{\theta}_0\right) \\
&= N\left(\left(\underset{\sim}{U}_m - \underset{\sim}{\theta}_0\right)' \Sigma_{\underset{\sim}{\theta}_0}^{-1} \left(\underset{\sim}{U}_m - \underset{\sim}{\theta}_0\right) + \mathrm{o}(1)\right) \\
&= N\left(\left(\underset{\sim}{U}_m - \underset{\sim}{\theta} + \underset{\sim}{\theta} - \underset{\sim}{\theta}_0\right)' \Sigma_{\underset{\sim}{\theta}_0}^{-1} \left(\underset{\sim}{U}_m - \underset{\sim}{\theta} + \underset{\sim}{\theta} - \underset{\sim}{\theta}_0\right) + \mathrm{o}(1)\right) \\
&= N\left(\left(\underset{\sim}{U}_m - \underset{\sim}{\theta}\right)' \Sigma_{\underset{\sim}{\theta}_0}^{-1} \left(\underset{\sim}{U}_m - \underset{\sim}{\theta}\right) + (\underset{\sim}{\theta} - \underset{\sim}{\theta}_0)' \Sigma_{\underset{\sim}{\theta}_0}^{-1} (\underset{\sim}{\theta} - \underset{\sim}{\theta}_0) + \mathrm{o}(1)\right) \\
&= N\left(\left(\underset{\sim}{\theta} - \underset{\sim}{\theta}_0\right)' \Sigma_{\underset{\sim}{\theta}}^{-1} \left(\underset{\sim}{\theta} - \underset{\sim}{\theta}_0\right) + \mathrm{o}(1)\right) \\
&\geq N\left(\frac{\| \underset{\sim}{\theta}_0 - \underset{\sim}{\theta} \|^2}{\omega_{\max}} + \mathrm{o}(1)\right) \to \infty, \qquad (9.26)
\end{aligned}$$

which implies the desired result.

Hence the proposed test is consistent.

Based on this result, I conjecture that for a sequence of contiguous alternatives $\underset{\sim}{\theta}^{(n_0)} = \frac{1}{K!} 1_m + \underset{\sim}{h}/\sqrt{n_0}$,

$$\chi_m^2 \xrightarrow{d} \chi^2\left(m, \lambda = \underset{\sim}{h}' \tilde{V}_{*m}^{-1} \underset{\sim}{h}\right), \qquad (9.27)$$

a non-central chi square distribution with m degrees of freedom and noncentrality parameter λ. I will investigate this issue in the future.

Chapter 10: Performance of the Rank Tests for Precedence Probabilities

In chapter 9, I introduced a test statistic in (9.3) for testing the hypothesis in (8.2) and discussed a few properties of the test. In this chapter, I report on a simulation study carried out to assess the performances of the proposed POT-type tests. To be a viable alternative to already existing standard tests, it is desirable that a new asymptotic test: (a) have type I error rates close to their nominal values for finite sample sizes; (b) be consistent; (c) be powerful for important classes of alternatives.

In Chapter 9, conditions were given under which tests for precedence are consistent. To illustrate the ability of the POT tests to have type I error rates close to their nominal values, I carried out the following simulations with $K = 3$ and $m = 6$.

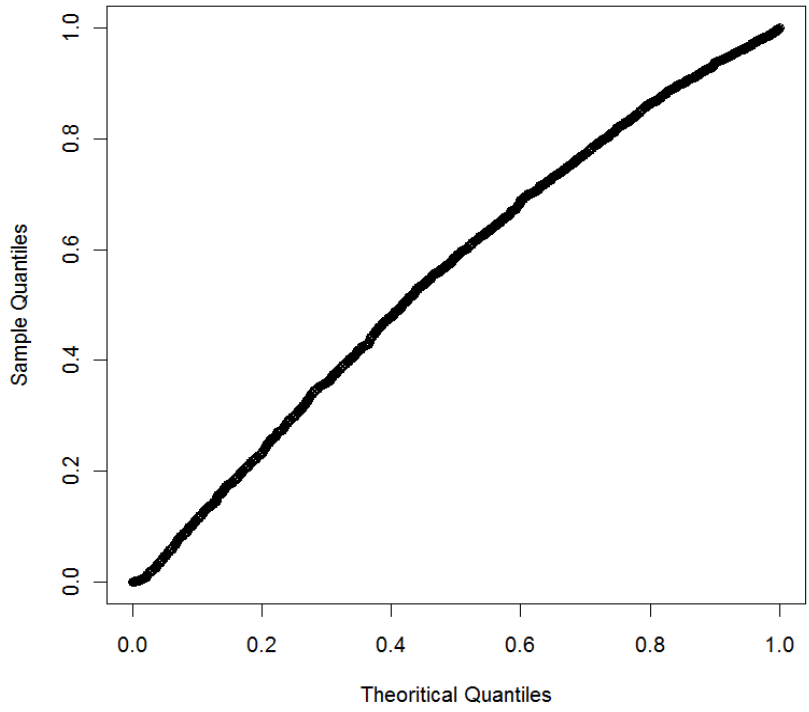
(a) Three independent samples each of size 30 was generated from a standard normal distribution. The proposed test statistic was calculated and a p-value obtained using its asymptotic chi square distribution.

This procedure was repeated 1000 times and a q-q plot of the resulting p-values versus (vs.) a Uniform $(0,1)$ distribution were plotted in Plot 10.1. The

plot is fairly linear and equiangular, is consistent with the test having approximately its nominal type I error rates in this case.

Plot 10.1: q-q plot of p-values vs. U(0,1) under H_0,

$$n_1 = n_2 = n_3 = 30, K = 3$$



For example, 5.3% of the 1000 values used Plot 10.1 are at most 0.05, such that a test run at the nominal type I error rate 0.05 would be judged to be performing satisfactorily. I carried out other simulations of this type (for which the results are not shown), for non-normal distributions, with sample sizes of 20, 50 and obtained similar favorable results. Additional support for the ability

of the POT to keep type I error close to a nominal 0.05 value is given in some of the simulated power tables presented below in the rows where all three distributions are equal.

Next, I report on simulations which illustrate the rates at which the tests' power functions approach 1 when the tests are carried at the nominal 0.05 rate.

(a) Independent random samples of equal size were generated from each of the underlying distributions and the test was performed and a p -value based on the test statistic's asymptotic distribution under the null hypothesis of homogeneity was calculated.

(b) This procedure was repeated 1000 times and the percentage of times for which p-value was at most 0.05 was tallied.

The rejection rate calculated in (b) above is used as an estimate for the power of the test. Table 10.1, which presents a few selected results, illustrates how the power goes to 1 for a variety of changes in location, scale and both location and scale for a variety of triples of normal distributions.

Table 10.1: Estimated Power of the POT for Different sample Sizes and Conditions

F_1	F_2	F_3	γ_m	γ_{sd}	$n_1 = n_2 = n_3$	Estimated Power
N(0,1)	N(0,1)	N(1,1.75)	0.154	0.292	10	0.401
					12	0.489
					15	0.641
					20	0.762
					25	0.869
					30	0.94
					35	0.972
					40	0.992
					50	0.995
					65	1.000
N(0,1)	N(1,1)	N(2,1)	0.443	0.486	10	0.887
					12	0.954
					15	0.99
					18	0.995
					20	1.000
N(0,1)	N(0,2)	N(0,3)	0.083	0.308	10	0.278
					12	0.367
					15	0.505
					20	0.697
					25	0.854
					30	0.947
					35	0.982
					40	0.995
					50	0.999
					55	1
N(0,1)	N(0.25,1.25)	N(0.75,1.5)	0.131	0.238	10	0.176
					12	0.221
					15	0.282
					20	0.357
					25	0.455
					30	0.571
					35	0.645
					40	0.726
					45	0.778
					50	0.838
					60	0.917
					70	0.952
					100	1

Additional simulations were carried out to assess how well the power of the new test compares to several well known competitors. In Chapter 4, the ANOVA F-test, Kruskal-Wallis (KW) Test and DISCO test by Rizzo and Szekely (2010) were described I compared the performance of the proposed test against these standard tests. As we will see, the power of the new test is not limited to changes just in locations. I begin by reviewing some standard tests for the K-sample problem.

First, the ANOVA F-test statistic is provided by,

$$F = \frac{\sum_{i=1}^{K}(n_i - 1)(\bar{X}_{i0} - \bar{X}_{00})^2 \Big/ (K-1)}{\sum_{i=1}^{K}\sum_{j=1}^{n_i}(X_{ij} - \bar{X}_{i0})^2 \Big/ (N-K)}, \qquad (10.1)$$

where $\bar{X}_{i0} = \sum_{j=1}^{n_i} X_{ij} \Big/ n_i$ and $\bar{X}_{00} = \sum_{i=1}^{K}\sum_{j=1}^{n_i} X_{ij} \Big/ N$.

The Kruskal-Wallis test statistic is given in (4.2).

Among many tests for scale parameters, I compare the proposed test with Levene's test, Bartlett's test, the Fligner-Killeen test and the Brown-Forsythe test.

Bartlett's (B) test is due to an idea proposed by Bartlett (1937) and first used by Snedecor and Cochran (1983). The test statistic is given by,

$$X_B^2 = \frac{(N-K)\ln(S_p^2) - \sum_{i=1}^{K}(n_i - 1)\ln(S_i^2)}{1 + \frac{1}{3(K-1)}\left(\sum_{i=1}^{K}\left(\frac{1}{n_i - 1}\right) - \frac{1}{N-K}\right)}, \tag{10.2}$$

where $S_p^2 = \frac{\sum_{i=1}^{K}(n_i - 1)S_i^2}{(N-K)}$ and $S_i^2 = \frac{\sum_{i=1}^{K}(X_{ij} - \bar{X}_{i0})^2}{(n_i - 1)}$.

I use a modified version of the Fligner-Killeen (FK) test, which was proposed in 1981 by Conover, Johnson and Johnson. Consider R_{ij}^* to be the rank of $|X_{ij} - median(X_{i1}, X_{i2}, ..., X_{in_i})|$ in the pooled sample and consider increasing scores $a_{N,l} = \Phi^{-1}\left(\frac{1 + \frac{l}{N+1}}{2}\right)$. The test statistic is given by,

$$X_{FK}^2 = \frac{\sum_{i=1}^{K} n_i(\bar{A}_i - \bar{a})^2}{\sum_{j=1}^{N}(a_{N,j} - \bar{a})^2/(N-1)}, \tag{10.3}$$

where $\bar{a} = \frac{1}{N}\sum_{j=1}^{N} a_{N,j}$ and $\bar{A}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} a_{N,R_{ij}^*}$

Levene's (1960) test (L) is probably the most-used of these tests. The test statistic is given by,

$$F_L = \frac{(N-K)\sum_{i=1}^{K} n_i(\bar{Z}_{i0} - \bar{Z}_{00})^2}{(K-1)\sum_{i=1}^{K}\sum_{j=1}^{n_i}(Z_{ij} - \bar{Z}_{i0})^2}, \tag{10.4}$$

where $Z_{ij} = | X_{ij} - \bar{X}_{i0} |$, $\bar{Z}_{i0} = \sum_{j=1}^{n_i} Z_{ij} \Big/ n_i$ and $\bar{Z}_{00} = \sum_{i=1}^{K}\sum_{j=1}^{n_i} Z_{ij} \Big/ N$.

The Brown-Forsythe (BF) test is a modification of Levene's test proposed by Brown and Forsythe (1974). The test statistic is

$$F_L = \frac{(N-K)\sum_{i=1}^{K} n_i(\bar{W}_{i0} - \bar{W}_{00})^2}{(K-1)\sum_{i=1}^{K}\sum_{j=1}^{n_i}(W_{ij} - \bar{W}_{i0})^2}, \tag{10.5}$$

where $W_{ij} = | X_{ij} - median(X_{i1}, X_{i2}, ..., X_{in_i}) |$, $\bar{W}_{i0} = \sum_{j=1}^{n_i} W_{ij} \Big/ n_i$ and $\bar{W}_{00} = \sum_{i=1}^{K}\sum_{j=1}^{n_i} Z_{ij} \Big/ N$.

To express the test statistic by Rizzo and Szekely, let us first define d_φ-distance between samples from two random variables X_i and X_j as, φ being a pre-determined constant between 0 and 2;

$$d_\varphi(X_i, X_j) = \frac{n_i n_j}{n_i + n_j}\left[2g_\varphi(X_i, X_j) - g_\varphi(X_i, X_i) - g_\varphi(X_j, X_j)\right], \tag{10.6}$$

where

$$g_\varphi(X_i, X_j) = \frac{1}{n_i n_j}\sum_{l=1}^{n_i}\sum_{m=1}^{n_j} || X_{il} - X_{jm} ||^\varphi . \tag{10.7}$$

Note that, in the expression above, $||.||$ denotes the Euclidean norm.

Then, let the between-sample dispersion be defined as,

$$S_{\varphi} = S_{\varphi}(X_1, X_2, ..., X_K) = \sum_{1 \le i < j \le K} \frac{n_i + n_j}{2N} d_{\varphi}(X_i, X_j). \quad (10.8)$$

Also, let the within-sample dispersion be defined as,

$$W_{\varphi} = W_{\varphi}(X_1, X_2, ..., X_K) = \sum_{1 \le i \le K} \frac{n_i}{2} g_{\varphi}(X_i, X_i). \quad (10.9)$$

The test statistic for testing the K-sample problem is provided by,

$$F_{\varphi} = \frac{S_{\varphi} / (K-1)}{W_{\varphi} / (N-K)}. \quad (10.10)$$

Note that the statistic depends on the choice of φ.

The POT was carried out with $m = K! - 1 = 5.$ To compare its power performance to those of the standard tests, I conducted the following simulation study.

(a) Independent random samples of equal size were generated from each of the underlying distributions and all the competing tests were carried out at a nominal type I error rate of 0.05 by deciding to reject the hypothesis of homogeneity if their p-values were at most 0.05.

(b) This procedure was repeated 1000 times and the percentage of times for which the null hypothesis was rejected was tallied for all the tests.

The rejection rates calculated in (b) above are estimated powers. Table 10.2(a) and 10.2(b) present a few of the selected results for normal and equal sample sizes of 10 and 20. Table 10.3(a) and 10.3(b) shows selected results under the

same conditions when the data are generated from a logistic distribution. In a similar manner, Table 10.4(a) and Table 10.4(b) list rejection rates, estimated powers, based on data generated from a Generalized Extreme Value distribution.

Table 10.2(a): Estimated Power of Different Tests for Normal Distributions with Sample size 10 from Each Distribution

F_1	F_2	F_3	ANOVA	KW	DISCO	FK	B	L	BS	POT
N(0,1)	N(0,1)	N(0,1)	0.056	0.05	0.05	0.031	0.046	0.047	0.031	0.058
N(0,1)	N(0,1)	N(.5,1.2)	0.137	0.139	0.149	0.046	0.101	0.095	0.055	0.15
N(0,1)	N(0,1)	N(.7,1.2)	0.253	0.262	0.261	0.046	0.101	0.095	0.055	0.233
N(0,1)	N(0,1)	N(.7,1.5)	0.193	0.215	0.244	0.116	0.214	0.208	0.124	0.254
N(0,1)	N(.2,1.2)	N(.7,1.5)	0.167	0.16	0.173	0.079	0.157	0.146	0.084	0.176
N(0,1)	N(.5,1.2)	N(.7,1.5)	0.205	0.18	0.166	0.079	0.157	0.146	0.084	0.167
N(0,1)	N(.2,1)	N(.5,1.5)	0.105	0.097	0.101	0.116	0.214	0.208	0.124	0.161
N(0,1)	N(0,1.2)	N(.5,1)	0.151	0.133	0.108	0.048	0.085	0.08	0.048	0.117
N(0,1)	N(1,1)	N(2,1)	0.957	0.954	0.926	0.041	0.047	0.069	0.04	0.897
N(0,1)	N(0,2)	N(0,3)	0.055	0.062	0.157	0.485	0.826	0.672	0.522	0.278
N(0,1)	N(0,1)	N(1,1.7)	0.258	0.292	0.354	0.207	0.404	0.365	0.242	0.401
N(0,1)	N(0,1)	N(2,2)	0.666	0.704	0.812	0.308	0.575	0.504	0.374	0.815
N(0,1)	N(0,1)	N(3,2.5)	0.817	0.865	0.949	0.512	0.826	0.717	0.609	0.95

Table 10.2(b): Estimated Power of Different Tests for Normal Distributions with Sample size 20 from Each Distribution

F_1	F_2	F_3	ANOVA	KW	DISCO	FK	B	L	BS	POT
N(0,1)	N(0,1)	N(0,1)	0.056	0.059	0.06	0.059	0.046	0.047	0.031	0.051
N(0,1)	N(0,1)	N(.5,1.5)	0.178	0.197	0.35	0.272	0.431	0.374	0.312	0.376
N(0,1)	N(0,1)	N(.7,1.2)	0.512	0.506	0.581	0.085	0.135	0.111	0.1	0.486
N(0,1)	N(0,1)	N(.7,1.5)	0.4	0.417	0.569	0.272	0.431	0.374	0.312	0.555
N(0,1)	N(.2,1.2)	N(.7,1.5)	0.329	0.317	0.405	0.182	0.306	0.26	0.21	0.357
N(0,1)	N(.5,1.2)	N(.7,1.5)	0.385	0.345	0.437	0.182	0.306	0.26	0.21	0.337
N(0,1)	N(.2,1)	N(.5,1.5)	0.167	0.165	0.313	0.272	0.431	0.374	0.312	0.341
N(0,1)	N(0,1.2)	N(.5,1)	0.296	0.269	0.301	0.109	0.158	0.163	0.129	0.259
N(0,1)	N(1,1)	N(2,1)	1	1	0.998	0.059	0.046	0.047	0.031	1
N(0,1)	N(0,2)	N(0,3)	0.05	0.053	0.542	0.92	0.993	0.967	0.949	0.697
N(0,1)	N(0,1)	N(1,1.7)	0.529	0.543	0.744	0.502	0.689	0.626	0.565	0.762
N(0,1)	N(0,1)	N(2,2)	0.953	0.956	1	0.686	0.868	0.809	0.755	0.992
N(0,1)	N(0,1)	N(3,2.5)	0.998	0.995	1	0.912	0.985	0.973	0.954	1

Table 10.3(a): Estimated Power of Different Tests for Logistic Distributions with Sample size 10 from Each Distribution

F_1	F_2	F_3	ANOVA	KW	DISCO	FK	B	L	BS	POT
L(0,1)	L(0,1)	L(0,1)	0.053	0.047	0.05	0.034	0.109	0.064	0.400	0.05
L(0,1)	L(1,1)	L(2,1)	0.533	0.546	0.524	0.034	0.109	0.064	0.400	0.488
L(0,1)	L(0,2)	L(0,3)	0.057	0.061	0.179	0.417	0.776	0.587	0.433	0.267
L(0,1)	L(1,2)	L(2,3)	0.168	0.173	0.328	0.417	0.776	0.587	0.433	0.296
L(0,1)	L(2,2)	L(4,3)	0.547	0.523	0.628	0.417	0.776	0.587	0.433	0.547
L(0,1)	L(2,1)	L(4,3)	0.697	0.664	0.862	0.596	0.891	0.795	0.691	0.886
L(0,2)	L(2,1)	L(4,3)	0.397	0.446	0.647	0.426	0.804	0.596	0.426	0.697
L(0,1)	L(0,1)	L(2,1)	0.648	0.667	0.66	0.034	0.109	0.064	0.400	0.586
L(0,1)	L(0,1)	L(2,2)	0.267	0.323	0.464	0.267	0.56	0.435	0.302	0.469
L(0,1)	L(0,2)	L(2,1)	0.568	0.515	0.596	0.244	0.532	0.412	0.295	0.574
L(0,1)	L(3,2)	L(6,3)	0.873	0.862	0.919	0.417	0.776	0.587	0.433	0.824
L(0,1)	L(3,2)	L(6,4)	0.773	0.725	0.898	0.624	0.926	0.788	0.671	0.819

Table 10.3(b): Estimated Power of Different Tests for Logistic Distributions with Sample size 20 from Each Distribution

F_1	F_2	F_3	ANOVA	KW	DISCO	FK	B	L	BS	POT
L(0,1)	L(0,1)	L(0,1)	0.048	0.053	0.058	0.048	0.138	0.060	0.049	0.058
L(0,1)	L(1,1)	L(2,1)	0.871	0.893	0.871	0.048	0.138	0.060	0.049	0.800
L(0,1)	L(0,2)	L(0,3)	0.047	0.058	0.467	0.892	0.982	0.937	0.901	0.64
L(0,1)	L(1,2)	L(2,3)	0.311	0.316	0.726	0.892	0.982	0.937	0.901	0.746
L(0,1)	L(2,2)	L(4,3)	0.890	0.878	0.937	0.892	0.982	0.937	0.901	0.939
L(0,1)	L(2,1)	L(4,3)	0.964	0.949	0.999	0.952	0.994	0.978	0.970	0.999
L(0,2)	L(2,1)	L(4,3)	0.717	0.784	0.973	0.876	0.975	0.931	0.890	0.977
L(0,1)	L(0,1)	L(2,1)	0.946	0.960	0.933	0.048	0.138	0.060	0.049	0.908
L(0,1)	L(0,1)	L(2,2)	0.519	0.613	0.820	0.646	0.86	0.766	0.725	0.883

Table 10.4(a): Estimated Power of Different Tests for Generalized Extreme Value Distributions with Sample size 10 from Each Distribution

F_1	F_2	F_3	ψ	F	KW	DISCO	FK	B	L	BS	POT
G(0,1)	G(0,1)	G(0,1)	0	0.052	0.047	0.056	0.044	0.153	0.111	0.043	0.05
G(0,1)	G(1,1)	G(2,1)	0	0.809	0.856	0.85	0.044	0.153	0.111	0.043	0.798
G(0,1)	G(0,2)	G(0,3)	0	0.088	0.084	0.214	0.407	0.77	0.604	0.404	0.243
G(0,1)	G(1,1)	G(2,1)	.5	0.407	0.671	0.603	0.102	0.653	0.294	0.044	0.766
G(0,1)	G(1,1)	G(2,1)	.2	0.648	0.777	0.769	0.067	0.363	0.165	0.046	0.772
G(0,1)	G(1,1)	G(2,1)	1	0.163	0.54	0.336	0.169	0.868	0.506	0.03	0.747
G(0,1)	G(1,1)	G(2,1)	-.5	0.978	0.98	0.981	0.036	0.055	0.08	0.04	0.935
G(0,1)	G(1,1)	G(2,1)	-.2	0.932	0.926	0.912	0.032	0.05	0.065	0.036	0.885
G(0,1)	G(0,2)	G(0,3)	-.2	.096	0.095	0.233	0.465	0.794	0.666	0.512	0.276
G(0,1)	G(0,2)	G(0,3)	-.5	0.100	0.105	0.263	0.468	0.792	0.667	0.485	0.331
G(0,1)	G(0,2)	G(0,3)	1	0.036	0.064	0.103	0.324	0.879	0.56	0.055	0.169
G(0,1)	G(0,2)	G(0,3)	.2	0.077	0.077	0.181	0.365	0.763	0.553	0.285	0.212
G(0,1)	G(0,2)	G(0,3)	.5	0.06	0.065	0.15	0.32	0.774	0.499	0.141	0.190
G(0,1)	G(1,2)	G(2,3)	.5	0.286	0.437	0.385	0.32	0.774	0.499	0.141	0.355
G(0,1)	G(1,2)	G(2,3)	0	0.613	0.536	0.639	0.407	0.77	0.604	0.404	0.514
G(0,1)	G(1,2)	G(2,3)	1	0.118	0.405	0.223	0.324	0.879	0.56	0.055	0.371

Table 10.4(b): Estimated Power of Different Tests for Generalized Extreme Value Distributions with Sample size 20 from Each Distribution

F_1	F_2	F_3	ψ	F	KW	DISCO	FK	B	L	BS	POT
G(0,1)	G(0,1)	G(0,1)	0	0.046	0.053	0.052	0.054	0.203	0.111	0.049	0.058
G(0,1)	G(1,1)	G(2,1)	0	0.994	0.997	0.999	0.054	0.203	0.111	0.049	0.992
G(0,1)	G(0,2)	G(0,3)	0	0.172	0.102	0.515	0.89	0.972	0.925	0.861	0.636
G(0,1)	G(1,1)	G(2,1)	.5	0.505	0.949	0.921	0.109	0.775	0.299	0.042	0.985
G(0,1)	G(1,1)	G(2,1)	.2	0.895	0.991	0.977	0.074	0.489	0.176	0.051	0.986
G(0,1)	G(1,1)	G(2,1)	1	0.16	0.86	0.568	0.178	0.928	0.487	0.026	0.981
G(0,1)	G(1,1)	G(2,1)	-.5	1	1	1	0.036	0.065	0.074	0.042	1
G(0,1)	G(0,2)	G(0,3)	-.5	0.123	0.162	0.749	0.936	0.994	0.974	0.95	0.749
G(0,1)	G(0,2)	G(0,3)	1	0.064	0.064	0.134	0.506	0.947	0.585	0.078	0.429
G(0,1)	G(0,2)	G(0,3)	.5	0.131	0.076	0.266	0.629	0.904	0.657	0.25	0.493
G(0,1)	G(1,2)	G(2,3)	.5	0.57	0.795	0.731	0.629	0.904	0.657	0.25	0.664
G(0,1)	G(1,2)	G(2,3)	0	0.959	0.893	0.952	0.89	0.972	0.925	0.861	0.873
G(0,1)	G(1,2)	G(2,3)	1	0.196	0.737	0.425	0.506	0.947	0.585	0.078	0.683

It is to be noted that, in the previous tables, the new test proposed here is denoted as POT, the ANOVA F test is denoted either as ANOVA or as F, the Kruskal-Wallis test is denoted as KW, Bartlett's test is denoted as B, the Fligner-Killeen test is denoted as FK, Levene's test is denoted as L, the Brown-

Forsythe test is denoted as BF and the Rizzo-Szekely test is denoted as DISCO while ψ denotes the shape parameter for the Generalized Extreme Value distribution.

As mentioned earlier, the calculation of the DISCO test statistic depends on φ. For the comparison purpose, I used the default $\varphi = 1$. I used the software *R* for calculation. To calculate p-values for the DISCO test, bootstrapping is needed. I used a bootstrap sample size of 250, as suggested by Rizzo and Szekely.

A few general observations:

i) The tests for detecting differences in scale parameters are mostly good in detecting differences in scales and are impervious to any change in location parameters. (noted exception is in the Generalized Extreme Value distributions)

ii) Bartlett's test cannot be trusted in non-normal cases, as the size of the test is much higher than what it aims for.

iii) Levene's test generally performs better than the Brown-Forsythe test.

iv) None of the tests considered here can detect the differences in the shape parameters for Generalized Extreme Value distributions.

v) The traditional ANOVA test cannot at all detect differences in the location parameters, if the shape parameter of the Generalized Extreme Value distribution is close to 1.

vi) Increasing the sample size seemed to increase the power of all the tests, when an increase in the power is expected.

Looking at the performance of the proposed test, we can conclude that:

(i) Its powers under the different considered alternatives approach 1.

(ii) It performs really well under a variety of conditions, changes in location, changes in scale and changes in both. The DISCO test proposed by Rizzo and Szekely (2010) is also good in this regard. But p-values for DISCO require a bootstrap that needs lots of CPU time.

(iii) The proposed test seems to be the best performer if the distribution with largest location parameter is also the one with the largest scale parameter.

In the next chapter, I check the performance of the proposed test for a single precedence probability.

Chapter 11: HUM and Related Problems

Let us again consider the motivating example discussed in Chapter 1. There are fourteen separate scales on which three groups of people are measured with a pre-decided scoring system. It is a well-established practice to judge the performance of medical classification procedures, such as these fourteen scales, using the *Hypervolume Under Manifolds* (HUM's) of their associated ROC manifolds. In Chapter 9, I described and explored a rank-based approach to provide a viable test for hypotheses in (8.2), which could be used to assess the performance of each scale based on asymptotically nonparametric tests about precedence probabilities. In this chapter, I focus on testing and interval estimation for a particular precedence probability being investigated for its ability to separate subjects into disease categories.

As described in Chapter 1, again consider a medical test designed to classify patients into one of $K \geq 2$ groups and let us suppose that the test employs a scoring system such that the score X_K from group K, is expected to be greater than a score from group $(K-1)$, X_{K-1} and so on. To test if data consisting of independent random samples support these expectations, I propose a particular test from the family of tests described in (8.2):

$$H_0 : F_1 = F_2 = ... = F_K ,$$

Vs.

$$H_a : P(X_1 < X_2 < ... < X_K) > 1/K!.$$

(11.1)

Here, we assume that X_i, a test score of a patient from group i, is an independent observation generated from an underlying, unspecified, continuous distribution function $F_i; i = 1, 2, ..., K$. Note that unlike (9.3), the alternative hypothesis here is explicitly one-sided. Since $1/4! = 0.04167$ and $1/5! = 0.0083$, the small values of the lower bounds in H_a in (11.1) are not very informative if H_0 is rejected and $K \geq 4$. A better way to assess how well the actual ordering conforms to the expected ordering would be to construct a confidence interval for $p(\xi_1) = P(X_1 < X_2 < ... < X_K) = U_1$, the HUM of the ROC manifold associated with this medical test.

11.1 A Test Statistic for HUM

In terms of Chapter 11, focusing on a case where the number of precedence probabilities being investigated, denoted by *m*, equals 1. Recall from Chapter 9 that for $K = 3$, the variance of U_1 under homogeneity is given by

$$\text{var}(U_1) = \frac{1}{n_1 n_2 n_3}\left(\frac{1}{6} + \left(\frac{n_1 + n_2 + n_3 - 3}{12}\right) + \left(\frac{(n_2 - 1)(n_1 + n_3 - 2)}{20}\right) + \left(\frac{(n_1 - 1)(n_3 - 1)}{30}\right)\right) - \frac{1}{36}. \quad (11.2)$$

Using asymptotic normality, I propose a new test statistic

$$Z_U = \frac{U_1 - (1/6)}{\sqrt{\text{var}(U_1)}} \quad (11.3)$$

for testing the hypotheses in (11.1). The null hypothesis is rejected at nominal type I error rate α, if and only if $Z_U > Z_\alpha$, where Z_α satisfies $P(Z > Z_\alpha) = \alpha$ and Z is a standard normal random variable. Apart from the tests proposed by Xiong *et al.* (2006), which were mentioned in Chapter 1, there are few other tests proposed for solving this problem. The test proposed by Nakas and Yiannoutsos (2004) merits special attention since they were the first to use the U-statistic approach in this type of problems. They proposed the following test statistic,

$$Z_{U^*} = \frac{U_1 - (1/K!)}{\hat{sd}(U_1)}, \quad (11.4)$$

where $\hat{sd}(U_1)$ is a data based estimate of the standard deviation of the statistic U_1. Nakas and Yiannoutsos proposed estimating the standard deviation using the bootstrap or using a component-wise *U*-statistic estimate for the standard

deviation. It could be argued that the test proposed in (11.3) is an improvement over the test in (11.4) since the variance does not have to be estimated.

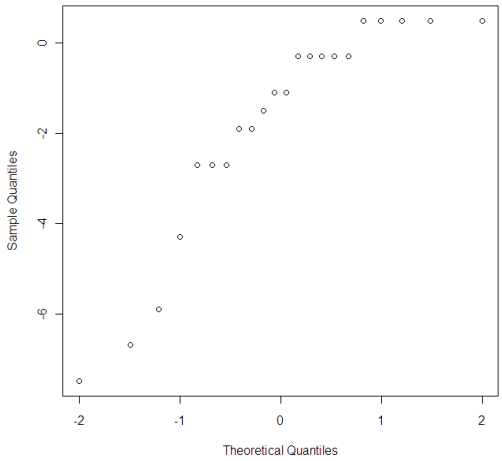
I compare below the performance of my new test with the performance of Xiong's test (2006) and the standard ANOVA F test. Xiong, *et al.* (2006) used the maximum likelihood approach to obtain a test statistic, V_{00}, the area under the ROC curve when $K = 2$. But, to create a test of significance, Xiong, *et al.* (2006) assumed all test scores to be either normally distributed or ones that could be monotonically transformed to a normal distribution, so that normal theory and likelihood based methods could be used.

The data Xiong used have 14 different clinical tests and each test was given to three a priori identified groups of people: diseased, healthy and transitional (those who are diagnosed with early stage AD). I obtained a copy of the data from Washington University Medical Center ADRC. While some of the test scores appeared to be normally distributed, some of them did not. Plot 11.1 presents, for example, a normal q-q plot for the test Visual Retention obtained from the diseased group of subjects, which suggests that at least some of the test scores are not normally distributed, a basic assumption for the test proposed by Xiong *et al.* (2006). The tests presented here have the advantage of not requiring that the scales be normally distributed.

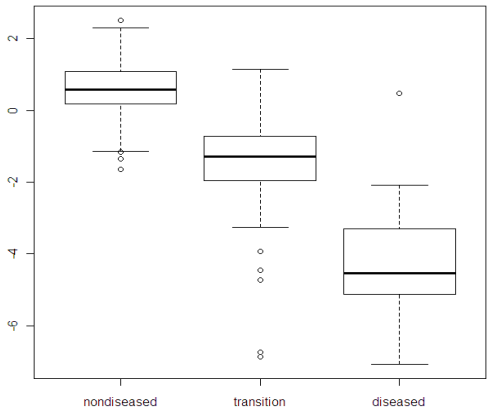
Now, taking a closer look at the data, there seems to be a pretty clear division among the three groups for all the scales. For example, side by side box plots of

the sores on “global factor” plotted in Plot 11.2, show a clear separation among the groups, a pattern typical for all fourteen scales.

Plot 11.1: Normal q-q plot for scores for people with AD in visual retention(copy)



Plot 11.2: Box plot for 3 groups of peoples’ test scores in Global factor



In fact, all of the scales indicate pretty conclusive separations among the three groups of subjects. Table 11.1 compares the results of my test applied to these data to the test proposed by Xiong, *et al.* (2006) and the traditional ANOVA F-test. Note that the traditional ANOVA test is a non-directional test while the other two tests are directional tests. To deal with the problem of tied observations, I used the following modified version of the proposed test statistic.

$$U_3 = \frac{1}{n_1 n_2 n_3}\Big(\sum_{i_1=1}^{n_1}\sum_{i_2=1}^{n_2}\sum_{i_3=1}^{n_3} I(X_{1i_1} < X_{2i_2} < X_{3i_3}) + \frac{1}{2}\sum_{i_1=1}^{n_1}\sum_{i_2=1}^{n_2}\sum_{i_3=1}^{n_3} I(X_{1i_1} = X_{2i_2} < X_{3i_3})$$

$$+\frac{1}{2}\sum_{i_1=1}^{n_1}\sum_{i_2=1}^{n_2}\sum_{i_3=1}^{n_3} I(X_{1i_1} < X_{2i_2} = X_{3i_3}) + \frac{1}{4}\sum_{i_1=1}^{n_1}\sum_{i_2=1}^{n_2}\sum_{i_3=1}^{n_3} I(X_{1i_1} = X_{2i_2} = X_{3i_3})\Big) \quad (11.5)$$

There are other ways of dealing with ties. For example, it could be argued that a medical test needs to clearly distinguish between any two groups so any tied data should not be counted in calculating the test statistic. In many data sets, medical data in particular, ties in the data can lead to incorrect conclusions. Unfortunately, dealing with tied observations is not a simple matter. The following tie-breaking methods could also be tried for obtaining a suitable test.

First, define the following values:

$$S_0 = \{(X_{1i_1}, X_{2i_2}, ..., X_{Ki_K}) : (X_{1i_1} \le X_{2i_2} \le ... \le X_{Ki_K}) \text{ with no ties}\}$$

$$S_1 = \{(X_{1i_1}, X_{2i_2}, ..., X_{Ki_K}) : (X_{1i_1} \le X_{2i_2} \le ... \le X_{Ki_K}) \text{ with } X_{li_l} = X_{(l+1)i_{(l+1)}} \text{ for 1 pair } (l, l+1)\}$$

$S_2 = \{(X_{1i_1}, X_{2i_2}, ..., X_{Ki_K}) : (X_{1i_1} \le X_{2i_2} \le ... \le X_{Ki_K})$ with $X_{li_l} = X_{(l+1)i_{(l+1)}}$ for 2 pairs $(l, l+1)\}$

.

.

$S_{K-1} = \{(X_{1i_1}, X_{2i_2}, ..., X_{Ki_K}) : (X_{1i_1} \le X_{2i_2} \le ... \le X_{Ki_K})$, $X_{li_l} = X_{(l+1)i_{(l+1)}}$ for $(K-1)$ pairs $(l, l+1)\}$.

Now, consider these following modified versions of the proposed test statistic:

$$U_K^{(h)}(\underline{Y}) = \frac{1}{n_1 n_2 ... n_K} \left\{ \sum_{j=0}^{K-1} \frac{1}{2^j} \sum_{S_j} I(X_{1i_1} \le X_{2i_2} \le ... \le X_{Ki_K}) \right\} \quad , \tag{11.6}$$

$$U_K^{(g)}(\underline{Y}) = \frac{1}{n_1 n_2 ... n_K} \left\{ \sum_{j=0}^{K-1} \frac{1}{(K!)^{j/(K-1)}} \sum_{S_j} I(X_{1i_1} \le X_{2i_2} \le ... \le X_{Ki_K}) \right\} , \tag{11.7}$$

$$U_K^{(w)}(\underline{Y}) = \frac{1}{2n_1 n_2 ... n_K} \left\{ \sum_{S_0} I(X_{1i_1} \le X_{2i_2} \le ... \le X_{Ki_K}) + \sum_{i_1=1}^{n_1} \sum_{i_2=1}^{n_2} ... \sum_{i_K=1}^{n_K} I(X_{1i_1} \le X_{2i_2} \le ... \le X_{Ki_K}) \right\} \tag{11.8}$$

Table 11.1 implies that the test statistic provided by both Xiong, *et al.* (2006) and the test statistic calculated by me are highly comparable, while the variance of my Z_U test statistic is a little smaller than the other test statistic, which is to be expected. In a few cases the standard deviation of the proposed test statistic is almost half of the estimated standard deviation of Xiong's proposed test statistic.

Looking at the Table 11.1, it is also evident that for all medical tests, the null hypothesis is rejected; i.e., the conclusion is that the HUM is significantly

higher than the average precedence probability under homogeneity. However, being larger than the average precedence probability does not necessarily mean that the HUM related to that medical test is the maximum among all precedence probabilities. Consider the following example.

Example 11.1: Let $X_1 \sim N(0,1)$, $X_2 \sim N(1.05,1)$, $X_3 \sim N(1,1)$.

Using the software *R* package "mnormt", the HUM is given by

$$p(\xi_1) = P(X_1 < X_2 < X_3) = 0.3127302 > 1/6 .$$

So, any decent test would hopefully reject the null hypothesis in (11.1) in this scenario. However, clearly, $P(X_1 < X_3 < X_2) = 0.3280731 > p(\xi_1)$. Although the null hypothesis in (11.1) should be rejected in this case, that does not automatically imply that the most common ordering (in this example, $X_1 < X_3 < X_2$) is the one anticipated, in this example, $X_1 < X_2 < X_3$.

So, instead of only doing hypothesis testing, we recommend constructing a confidence interval for HUM. We explore this idea in the next section of this chapter.

Table 11.1: P-value for testing hypothesis (11.1) for different factors in AD

Factor (Medical Test)	U_1	Asymptotic s.d.	P value	ANOVA F	P value	V_{00}	Estimated s.d.	P value
Global Factor	0.7747016	0.04169711	0.000	25.3999	0.000	0.728	0.05255102	0.000
Temporal Factor	0.7710744	0.04104448	0.000	56.9518	0.000	0.752	0.05102041	0.000
Parietal Factor	0.5983389	0.04169711	0.000	40.8876	0.000	0.555	0.0622449	0.000
Frontal Factor	0.6565276	0.04169711	0.000	75.9394	0.000	0.657	0.05408163	0.000
Logical Memory	0.7374632	0.0399638	0.000	16.8734	0.000	0.724	0.04744898	0.000
Digit Span Forward	0.5165987	0.03902601	0.000	11.7316	3.746 e-05	0.522	0.05561224	0.000
Digit Span backward	0.6072836	0.04001238	0.000	11.0928	5.92e-05	0.599	0.05612245	0.000
Informat-ion	0.6924388	0.03902601	0.000	13.4796	9.637 e-06	0.680	0.05102041	0.000
Visual Retention (10s)	0.5888028	0.04169711	0.000	16.4109	1.177 e-06	0.587	0.05357143	0.000
Visual Retention (copy)	0.3707645	0.04104448	3.3e-07	2.7323	0.079	0.347	0.05102041	2e-04
Boston Naming	0.6323384	0.03947893	0.000	12.5157	2.061 e-05	0.573	0.05561224	0.000
Mental Control	0.5069929	0.04001238	0.000	13.9053	7.105 e-06	0.532	0.05357143	0.000
Word Fluency	0.5801309	0.04104448	0.000	5.8184	0.004 446	0.568	0.05306122	0.000
Associate Learning	0.6720863	0.04118809	0.000	10.0851	0.000 1303	0.630	0.05408163	0.000

11.2 Bootstrap Confidence Interval for Precedence Probability

As mentioned in the previous section, constructing a confidence interval for a precedence probability, particularly for HUM, might be more informative than just reporting a p-value for the test in (11.1). Xiong, *et al.* (2006) used the assumption of normality to obtain a confidence interval for HUM. However, a confidence interval based on an assumption of normality can perform poorly if that assumption is not correct, as is evident from Plot 11.1. Instead of a normal-based procedure, I propose the following simple, bootstrap percentile confidence interval to obtain an approximate $100(1-\alpha)\%$ confidence interval for any specified precedence probability. Efron (1979) coined the term bootstrap and introduced this particular version of resampling to the science of statistics. Note that (11.3) cannot be used to construct a confidence interval for HUM since it is constructed under the assumption of homogeneity.

Without loss of generality, we construct a confidence interval for the particular HUM= $p(\xi_1)$. Let $U_1(\mathbf{X})$ be the U-Statistic estimate of $p(\xi_1)$ based on independent pooled N dimensional data vector $\mathbf{X} = \{X_{ij}\}$.

(A) Let $\hat{F}_i$ be an estimate (empirical CDF) of F_i obtained from $\{X_{ij}; j = 1,2,...,n_i\}$ and set $\hat{\underline{\mathbf{F}}} = \{\hat{F}_i, i = 1,2,...,K\}$, $\underline{\mathbf{F}} = \{F_i\}$.

(B) Let $\underline{\mathbf{y}}^* = \{\underline{\mathbf{y}}_1^*, \underline{\mathbf{y}}_2^*, ..., \underline{\mathbf{y}}_M^*\}$ be M independent bootstrap samples, $\underline{\mathbf{y}}_j^* = \{\underline{\mathbf{y}}_{ji}^* = \{y_{jil}^*, l = 1, 2, ..., n_i\}$ drawn from $\hat{F}_i; i = 1, 2, ..., K\}\}$, $j = 1, 2, ..., M$.

(C) Obtain $U_1(\underline{\mathbf{y}}_j^*)$; $j = 1, 2, ..., M$. Let, $U_1(\underline{\mathbf{y}}_{(1)}^*) \le U_1(\underline{\mathbf{y}}_{(2)}^*) \le ... \le U_1(\underline{\mathbf{y}}_{(M)}^*)$.

(D) Then, a percentile $100(1-\alpha)\%$ confidence interval for $p(\xi_1)$ is given by

$$\left(U_1(\underline{\mathbf{y}}_{(LL)}^*), U_1(\underline{\mathbf{y}}_{(UL)}^*)\right); \tag{11.9}$$

where $LL = \max\left[1, \text{int}(M \times \alpha / 2)\right]$, $UL = \min\left[M, \{\text{int}(M \times ((1-\alpha)/2) + 1\}\right]$ and $\text{int}(a)$ denotes the largest integer smaller than a real, positive number a.

One of the advantages of using this interval instead of the usual bootstrap confidence interval is that, even if $p(\xi_1)$ is close to either 0 or 1, this percentile interval will always be within $[0,1]$. Note that the usual bootstrap intervals would have been expressed by

$$\left(2U_1(\underline{\mathbf{y}}) - U_1(\underline{\mathbf{y}}_{(UL)}^*), 2U_1(\underline{\mathbf{y}}) - U_1(\underline{\mathbf{y}}_{(LL)}^*)\right). \tag{11.10}$$

I used simulation to investigate the average width and coverage rate of these percentile confidence intervals under different situations with $1-\alpha$ = 0.95. A few of my results are presented in Table 11.2. From this study it seems that the width of the confidence intervals, as expected from rank-based inference, does not depend on the underlying distributions. For this study, first I generated a random sample of size n; $n = 20, 30, 50$ from each of the underlying

distributions. A bootstrap sample of size 1000 was considered and the bootstrap percentile confidence intervals were calculated. This procedure was repeated 200 times to obtain the estimated coverage rate and average width of these confidence intervals. The average width, as expected, appears to depend inversely on n, larger sample sizes leading to shorter intervals. The estimated coverage rates tend to be slightly higher than their nominal value of 0.95.

Table 11.2: Average Width and Coverage Rate of .95 Percentile Bootstrap Confidence Intervals

F_1	F_2	F_3	$p(\xi_1)$	n	M	Average Width	Coverage Rate
N(0,1)	N(0,1)	N(0,1)	0.1667	50	1000	0.1281	0.97
				30	1000	0.1527	0.965
N(0,1)	N(1,1)	N(2,1)	0.5362	50	1000	0.1836	0.985
N(0,1)	N(2,1)	N(4,1)	0.8430	50	1000	0.1295	0.96
N(0,1)	N(3,1)	N(6,1)	0.9661	50	1000	0.0521	0.965
Exp(1)	Exp(1)	Exp(1)	0.1667	30	1000	0.1584	0.985
				20	1000	0.1953	1.000
Exp(3)	Exp(2)	Exp(1)	0.3333	20	1000	0.2774	0.995
				30	1000	0.2226	0.955
				50	1000	0.1800	0.98
Exp(9)	Exp(5)	Exp(1)	0.5	50	1000	0.1959	0.995
				30	1000	0.2556	0.96
				20	1000	0.2852	0.995

I then decided to compare the performance of this bootstrap confidence interval to the normal theory confidence interval given by Xiong, *et al.* (2006). I used a bootstrap sample size 250 for this study and looked at 95% confidence intervals. The results are summarized in Table 11.3.

Table 11.3: Confidence Intervals for HUM in AD

Factor (Medical Test)	U_1	Bootstrap Confidence Interval	V_{00}	Estimated s.d.	Xiong's Confidence Interval
Global Factor	0.7747016	(0.657,0.884)	0.728	0.05255102	(0.625,0.831)
Temporal Factor	0.7710744	(0.661,0.870)	0.752	0.05102041	(0.652,0.852)
Parietal Factor	0.5983389	(0.466,0.724)	0.555	0.0622449	(0.433, 0.676)
Frontal Factor	0.6565276	(0.543,0.767)	0.657	0.05408163	(0.551,0.763)
Logical Memory	0.7374632	(0.637,0.837)	0.724	0.04744898	(0.631,0.818)
Digit Span Forward	0.5165987	(0.408,0.632)	0.522	0.05561224	(0.413,0.631)
Digit Span backward	0.6072836	(0.494,0.722)	0.599	0.05612245	(0.489,0.708)
Information	0.6924388	(0.590,0.793)	0.680	0.05102041	(0.580,0.779)
Visual Retention(10s)	0.5888028	(0.464,0.708)	0.587	0.05357143	(0.482,0.691)
Visual Retention(copy)	0.3707645	(0.285,0.470)	0.347	0.05102041	(0.247,0.447)
Boston Naming	0.6323384	(0.516,0.743)	0.573	0.05561224	(0.464,0.682)
Mental Control	0.5069929	(0.400,0.628)	0.532	0.05357143	(0.427,0.637)
Word Fluency	0.5801309	(0.456,0.695)	0.568	0.05306122	(0.464,0.672)
Associate Learning	0.6720863	(0.556,0.790)	0.630	0.05408163	(0.524,0.736)

From the table, the confidence intervals seem to be of comparable width. Next, to assess the performance of Xiong's normal-theory confidence intervals when normality does not hold, I conducted a small-scale simulation study using data from a Cauchy distribution, a distribution having much heavier tails than the normal distribution. If X is a random variable that follows a Cauchy distribution with location parameter μ and scale parameter σ, its distribution function is given by

$$F(x) = \frac{1}{\Pi}\arctan\left(\frac{x-\mu}{\sigma}\right)+\frac{1}{2} \ . \qquad (11.11)$$

A simulation study by me indicates that when the underlying distribution is Cauchy, and all 3 distributions are identical, Xiong's method overestimates HUM, which leads to low coverage rates for his confidence intervals. However, the bootstrap percentile confidence interval works well here, as was true for the non-normal distributions in Table 11.2, with coverage rates close to nominal.

In Table 11.3, a very interesting confidence interval is the one for the HUM for the "Visual retention (copy)" factor. The lower confidence limits for this particular HUM, using both these methods, are well below 0.30. This leads one to consider a possible situation such as Example 11.1; where a precedence probability, although significantly greater than the average precedence probability was shown not to be the maximum of precedence probabilities. I investigate this matter in the next section of this chapter.

11.3 Comparing Precedence Probabilities

In Example 11.1, I noted that just testing hypothesis (11.1) might not always be a good approach to checking the performance of a medical test used to separate groups. A good medical test that separates groups according to an expected ordering such as $X_1 < X_2 < \cdots < X_K$ should result in the largest precedence probability being assigned to this ordering. Accordingly, in such cases, I propose testing

$$H_0 : F_1 = F_2 = ... = F_K$$

Vs.

$$H_A : p(\xi_1) > p(\xi_i); i = 2,3,...,K! \tag{11.12}$$

However, constructing a test for such a hypothesis is very difficult. Instead, suppose that the researcher has a prior belief that another precedence probability, say $p(\xi_2)$, might be the closest in value to $p(\xi_1)$. In that case, the following hypotheses could then be of interest:

$$H_0 : F_1 = F_2 = ... = F_K$$

Vs.

$$H_A : p(\xi_1) > p(\xi_2). \tag{11.13}$$

I propose the following method for implementing this test.

i) Calculate a bootstrap confidence interval for $p(\xi_1)$, as described in Section 11.2.

ii) If a lower limit of the confidence interval in i) is more than 0.5, reject the null hypothesis in (11.13).

iii) If the lower limit of the confidence interval in (i) is less than 0.5, calculate a bootstrap confidence interval for $p(\xi_1) - p(\xi_2)$ in the following manner:

Let $U_i(\underline{Y})$ be the U-Statistic estimate of $p(\xi_i); i = 1,2$ based on independent pooled N dimensional data vector $\underline{Y} = \{X_{ij}\}$.

(A) Let $\hat{F}_i$ be an estimate (empirical CDF) of F_i obtained from $\{X_{ij}; j = 1,2,...,n_i\}$ and set $\underline{\hat{\mathbf{F}}} = \{\hat{F}_i, i = 1,2,...,K\}$, $\underline{\mathbf{F}} = \{F_i\}$.

(B) Let $\underline{\mathbf{y}}^* = \{\underline{\mathbf{y}}_1^*, \underline{\mathbf{y}}_2^*, ..., \underline{\mathbf{y}}_M^*\}$ be M independent bootstrap samples, $\underline{\mathbf{y}}_j^* = \{\underline{\mathbf{y}}_{ji}^* = \{y_{jil}^*, l = 1,2,...,n_i\}$ drawn from $\hat{F}_i; i = 1,2,...,K\}\}$, $j = 1,2,...,M$.

(C) Obtain $U_{12}(\underline{\mathbf{y}}_j^*) = U_1(\underline{\mathbf{y}}_j^*) - U_2(\underline{\mathbf{y}}_j^*)$; $j = 1,2,...,M$.

(D) Let, $U_{12}(\underline{\mathbf{y}}_{(1)}^*) \le U_{12}(\underline{\mathbf{y}}_{(2)}^*) \le ... \le U_{12}(\underline{\mathbf{y}}_{(M)}^*)$.

(E) Then a $100(1-\alpha)\%$ confidence interval for $p(\xi_1)$ - $p(\xi_2)$ is given by

$$\left(U_{12}(\underline{\mathbf{y}}_{(LL)}^*), U_{12}(\underline{\mathbf{y}}_{(UL)}^*)\right); \tag{11.14}$$

where $LL = \max\left[1, \text{int}(M \times \alpha / 2)\right]$, $UL = \min\left[M, \{\text{int}(M \times ((1-\alpha)/2) + 1\}\right]$ and $\text{int}(a)$ is the largest integer smaller than a real number a.

iv) If the lower limit of the confidence interval in (11.14) is greater than 0, reject the null hypothesis in (11.13).

Consider our motivating example. In the case of Alzheimer's disease, it is often more difficult to detect the onset of the disease. So, a *not-so-good* medical test, might have a problem in separating healthy people from early-stage Alzheimer's patients. Let X_1, X_2 and X_3 denote independent values of a continuous score T obtained by subjects from healthy, intermediate and diseased groups, respectively. Then, for a not-so-good medical test, the probability of the expected precedence ordering of $\{X_1 < X_2 < X_3\}$ might not be larger than $P\{X_2 < X_1 < X_3\}$. From this perspective, looking at Table 11.3, we note that for 7 out of the 14 medical tests considered, the bootstrap percentile confidence interval's lower limit was less than 0.5. Please note that, if $p(\xi_1) > 0.5$, that means $p(\xi_1) > p(\xi_i); i = 2,3,...,K!$.

However, looking at the bootstrap confidence intervals for $p(\xi_1) - p(\xi_2)$, only one of those 14 tests had a confidence interval with lower limit smaller than 0. The "visual retention (copy)" factor was found to have a bootstrap confidence interval $[-0.007, 0.139]$ based on a bootstrap sample of size 1000. This possibly indicates that "visual retention (copy)" is not a very good test for separating all three groups of people. It might be effective in discriminating between healthy people and diseased people. But, that was not the intention of this medical test.

Although, Xiong, *et al.* (2006) concluded that all the 14 medical tests are effective in detecting all three groups of people, based on my analysis, using this new approach, I would not recommend "visual retention (copy)" as an effective test for properly separating all three groups of patients.

Chapter 12: Future Research and Summary

Recall that throughout my proposal $\{X_i \sim F_i; i = 1,2,...,K\}$ are jointly independent random variables and the distribution functions $\{F_i\}$ are assumed to be continuous. Also consider the rank vectors $\mathbf{R} = \{\underline{R}_i;\ i = 1,2,...,K\}$, where the components of $\underline{R}_i$ are the ranks of $\{X_{ij}, j = 1,2,...,n_i\}$ in the pooled N dimensional data vector $\mathbf{X} = \{X_{ij}\}$.

[1] In Chapters 6 and 7, I constructed and described three indices for measuring the distances among two or more distributions. This seems to be a very promising idea and I have already started new work on this subject. Recall that the three index measures, all based on precedence probabilities, are defined as;

$$\gamma_m = \frac{\max_i \left(p(\xi_i), i = 1,2,...,K! \right) - \frac{1}{K!}}{1 - \frac{1}{K!}}, \tag{12.1}$$

$$\gamma_{ss} = \frac{\sum_{i=1}^{K!} p(\xi_i)^2 - \frac{1}{K!}}{1 - \frac{1}{K!}}, \tag{12.2}$$

$$\gamma_{sd} = \frac{s.d.\{p(\xi_i); i = 1,2,...,K!\}}{\sqrt{\left(\frac{1}{K!}\right)}}. \tag{12.3}$$

In this dissertation I used U-statistic estimates for precedence probabilities for testing a family of hypotheses. This naturally leads to basing inference for these three index measures on a U-statistic approach . Specifically, if U_i is the U-statistic estimator of precedence probability $p(\xi_i) \forall i = 1,2,...,K!$, I propose the following estimates of distance index measures.

$$\hat{\gamma}_m = \frac{\max_i \left(U_i, i = 1,2,...,K!\right) - \frac{1}{K!}}{1 - \frac{1}{K!}}, \tag{12.4}$$

$$\hat{\gamma}_{ss} = \frac{\sum_{i=1}^{K!} U_i^2 - \frac{1}{K!}}{1 - \frac{1}{K!}}, \tag{12.5}$$

$$\hat{\gamma}_{sd} = \frac{s.d.\{U_i; i = 1,2,...,K!\}}{\sqrt{\left(\frac{1}{K!}\right)}}. \tag{12.6}$$

I intend to check the characteristics and performances of these estimates in the near future. It would also be interesting to obtain confidence intervals for these index measures. Since estimating the standard errors of these statistics would require knowledge of the parametric forms of the distributions and could be difficult or impossible to calculate, I propose using bootstrapping to achieve this goal.

[2] In many data sets and medical data in particular, ties among the observations violate a key assumption and can result in incorrect inferences. Dealing with tied observations in the setting of precedence probabilities is an important and difficult issue. I plan to explore different tie-breaking methods for the *U* statistics I employed. To help the reader recall the methods I proposed in Chapter 9 for tie breaking, again define

$S_0 = \{(X_{1i_1}, X_{2i_2}, ..., X_{Ki_K}) : (X_{1i_1} \le X_{2i_2} \le ... \le X_{Ki_K})$ with no ties},

$S_1 = \{(X_{1i_1}, X_{2i_2}, ..., X_{Ki_K}) : (X_{1i_1} \le X_{2i_2} \le ... \le X_{Ki_K})$ with $X_{li_l} = X_{(l+1)i_{(l+1)}}$ for 1 pair $(l, l+1)\}$,

$S_2 = \{(X_{1i_1}, X_{2i_2}, ..., X_{Ki_K}) : (X_{1i_1} \le X_{2i_2} \le ... \le X_{Ki_K})$ with $X_{li_l} = X_{(l+1)i_{(l+1)}}$ for 2 pairs $(l, l+1)\}$,

.

.

$S_{K-1} = \{(X_{1i_1}, X_{2i_2}, ..., X_{Ki_K}) : (X_{1i_1} \le X_{2i_2} \le ... \le X_{Ki_K})$, $X_{li_l} = X_{(l+1)i_{(l+1)}}$ for $(K-1)$ pairs $(l, l+1)\}$.

Then, consider the following modified versions of the proposed test statistic.

$$U_K^{(h)}(\underline{Y}) = \frac{1}{n_1 n_2 ... n_K} \left\{ \sum_{j=0}^{K-1} \frac{1}{2^j} \sum_{S_j} I(X_{1i_1} \le X_{2i_2} \le ... \le X_{Ki_K}) \right\} \quad (12.7)$$

$$U_K^{(g)}(\underline{Y}) = \frac{1}{n_1 n_2 ... n_K} \left\{ \sum_{j=0}^{K-1} \frac{1}{(K!)^{j/(K-1)}} \sum_{S_j} I(X_{1i_1} \le X_{2i_2} \le ... \le X_{Ki_K}) \right\} \quad (12.8)$$

$$U_K^{(w)}(\underline{Y}) = \frac{1}{2n_1 n_2 ... n_K} \left\{ \sum_{S_0} I(X_{1i_1} \le X_{2i_2} \le ... \le X_{Ki_K}) + \sum_{i_1=1}^{n_1} \sum_{i_2=1}^{n_2} ... \sum_{i_K=1}^{n_K} I(X_{1i_1} \le X_{2i_2} \le ... \le X_{Ki_K}) \right\} \quad (12.9)$$

Depending on circumstances, any one of these tie-breaking methods might be more suitable than the others. It is not intuitively clear how to select any one of these methods over the others for all purposes without properly exploring their behavior. I intend to investigate all three modified versions of the proposed test statistic to check their performance and their asymptotic behavior.

[3] Although the bootstrap percentile confidence interval seems to work well for any precedence probability, I intend to study pre-pivoted bootstrap confidence intervals to see if it yields shorter confidence intervals with satisfactory coverage rates. Specifically, constructing a confidence interval estimate of $p(\xi_1)$ of the form [*L, U*] or [*L*, 1) or $(0,U]$ would be of great practical interest, especially if *L* were close to 1 or *U* close to 0. The following bootstrap procedure can be used to accomplish this. Simulation would then be used to assess the performance of this procedure in terms of coverage rate and width. The algorithm is inspired by Beran (1987).

Let $U_1(\underline{Y})$ be the U-Statistic estimate of $p(\xi_1)$ based on independent pooled N dimensional data vector $\underline{Y} = \{X_{ij}\}$. Let $\hat{F}_i$ be an estimate (parametric or empirical CDF) of F_i obtained from $\{X_{ij}; j = 1,2,...,n_i\}$, and set $\underline{\hat{\mathbf{F}}} = \{\hat{F}_i, i = 1,2,...,K\}$, $\underline{\mathbf{F}} = \{F_i\}$. Then, we have $p(\xi_1) = p(\underline{\mathbf{F}})$ and let $U_1(\underline{Y}) = p(\underline{\hat{\mathbf{F}}})$. Pre-pivoting is based on estimating the distribution of the root $R = p(\underline{\hat{\mathbf{F}}}) - p(\underline{\mathbf{F}})$.

I describe below a suggested algorithm for a bootstrap confidence interval for $p(\xi_1)$.

(A) Let $\underline{\mathbf{y}}^* = \{\underline{\mathbf{y}}_1^*, \underline{\mathbf{y}}_2^*, ..., \underline{\mathbf{y}}_M^*\}$ be M independent bootstrap samples, $\underline{\mathbf{y}}_j^* = \{\underline{\mathbf{y}}_{ji}^* = \{y_{jil}^*, l = 1,2,...,n_i\}$ drawn from $\hat{F}_i$; $i = 1,2,...,K\}\}$, $j = 1,2,...,M$. Let $\{\underline{\hat{\mathrm{F}}}_j^* = \{\hat{F}_{ji}^*; i = 1,2,...,K\}$ be estimates of $\{\hat{F}_i\}$, obtained from $\underline{\mathbf{y}}_{ji}^*$, $i = 1,2,...,K$, $j = 1,2,...,M$. Compute $\hat{p}(\{\underline{\hat{\mathbf{F}}}_j^*\})$ and the corresponding *roots* $R_j = p(\hat{\mathbf{F}}_j^*) - p(\hat{\mathbf{F}})$, $j = 1,2,...,M$. For large *M*, let $\hat{H}(r, \underline{\hat{\mathbf{F}}}) = \#\{\mathrm{j};\ \mathrm{R_j} < r\}/M$ be the left continuous, empirical CDF obtained from $\{R_j : 1 \le j \le M\}$.

(B) Let $\{\underline{\mathbf{y}}_{j,1}^{**}, \underline{\mathbf{y}}_{j,2}^{**}, ..., \underline{\mathbf{y}}_{j,B}^{**}\}$ be B, independent bootstrap random samples from $\hat{\mathbf{F}}_j^*$ and compute $\hat{\mathbf{F}}_{jk}^{**}$ and $\hat{p}(\hat{\mathbf{F}}_{j,k}^{**})$ from $\underline{\mathbf{y}}_{j,k}^{**}$ and $R_{j,k} = p(\hat{\mathbf{F}}_{j,k}^{**}) - p(\underline{\hat{\mathbf{F}}}_j^*)$, for $j = 1,2,...,M$, $k = 1,2,...,B$.

(C) Compute $Z_j = \dfrac{\#\{k;\ R_{j,k} \le R_j\}}{B}$, for $j = 1,2,...,M$. Let $\hat{H}_1(z, \underline{\hat{\mathbf{F}}}^*) = \#\{j; Z_j < z\}/M$ be the left continuous empirical CDF obtained from $\{Z_j : 1 \le j \le M\}$.

(D) Then, an approximate, lower, $1-\alpha$ confidence interval p is given by $[L, 1)$, where

$$L = \hat{H}^{-1}(\hat{H}_1^{-1}(1-\alpha)) \qquad (12.10)$$

I believe this algorithm could lead to an improvement over the much simpler percentile bootstrap confidence interval I used in Chapter 11 and intend to explore it in the future.

[4] Recall from Chapter 11 the interesting problem of checking whether a particular precedence has the maximum probability among all or a class of orderings. Specifically, consider our motivating example of Alzheimer's patients from Chapter 1. Usually, to check the performance a clinical test, the following hypothesis is tested.

$$H_0 : F_1 = F_2 = ... = F_K$$

Vs.

$$H_A : p(\xi_1) > \frac{1}{K!} \tag{12.11}$$

As argued in Chapter 11, a better hypothesis to test in the above situation would be

$$H_0 : F_1 = F_2 = ... = F_K$$

Vs.

$$H_A : p(\xi_1) > p(\xi_i); i = 2,3,...,K! \tag{12.12}$$

Of course, it is not easy to obtain a test statistic and conduct the test mentioned in (12.12). One way to perform the hypothesis test would be to use a simultaneous confidence interval approach based on a bootstrap. Simultaneous bootstrap confidence intervals for $\{p(\xi_1) - p(\xi_i); i = 2,3,...,K!\}$ can

be obtained using a method proposed by Mandel and Betensky (2008). I plan to use the U-statistic estimate as employed in Chapter 11 to investigate this approach. I propose rejecting the null hypothesis in (12.12) if the lower limits of all the confidence intervals are greater than zero. The scope and performance of this approach would be of interest to many researchers, particularly in the medical field.

[5] For $K=3$, given the random sample $\{X_{ij}; i=1,2,3; j=1,2,..,n_i\}$, consider $U_{123} = \#\{X_{1i_1} < X_{2i_2} < X_{3i_3}\}$ to be the precedence ordering U-Statistic and let the pair-wise Mann-Whitney statistics be given by $U_{12} = \#\{X_{1i_1} < X_{2i_2}\}$, $U_{13} = \#\{X_{1i_1} < X_{3i_3}\}$ and $U_{23} = \#\{X_{2i_2} < X_{3i_3}\}$.

Then, define

$$U^*_{123} = h(U_{123}, U_{12}, U_{13}, U_{23}), \tag{12.13}$$

where h is increasing in each of its arguments. Whitney (1951) used a special case of this approach. For appropriate h, U^*_{123} could be used as the basis of either a non-directional or directional alternative. I plan to explore this type of statistic in more detail in the future.

[6] Recall that in addition to using the asymptotic chi-square distribution to obtain p-values for an observed value of $\chi^2_{m(K)}$ given in (9.3), I proposed an algorithm for computing a permutation p-value in Chapter 9. Although computationally a little bit more time-consuming, this approach provides another option for carrying out a truly nonparametric test for the hypothesis in (8.2). I plan to study this approach in the future and compare it to the asymptotic nonparametric test discussed in detail in this dissertation.

In summary, my research involves the study of stochastic precedence and precedence probabilities, which are currently being used in many fields, especially the medical sciences, mostly for only two distributions. In this thesis, for $K \geq 2,$ I proposed asymptotically non-parametric, rank-based test statistic which is also based on *U*-statistic estimators of precedence probabilities for a family of hypothesis tests which includes the very well known K-sample problem as well as the problem of testing for HUM of a *ROC* manifold. This test statistic performs well under various conditions, including sensitivity to differences in locations, scales or both, is easy to compute, and provides a viable alternative to the existing standard tests. In this thesis, I also proposed new distance index measures among distributions based on precedence probabilities. These measures have the potential to be important, new "effect sizes", especially for research based on clinical trials. Hopefully, my research will expand and significantly contribute to nonparametric and asymptotically

nonparametric inference for comparing more than two distributions, which has been and will remain an important area for research.

References

Anderson, T. W. (1962) "On the Distribution of the Two-sample Cramer-von Mises Criterion," *The Annals of Mathematical Statistics*, **33**, 1148-1159.

Arcones, M. A., Kvam, P. H. and Samaniego, F. J. (2002) "Nonparametric Estimation of a Distribution Subject to a Stochastic Precedence Constraint," *Journal of the American Statistical association*, **97**, 170-182.

Bartlett, M. S. (1937) "Properties of Sufficiency and Statistical Tests," *Proceedings of the Royal Statistical Society Series A*, **160**, 268-282.

Behnen, K. and Neuhaus, G., *Rank Tests with Estimated Scores and Their Application*, B. G. Teubner, Stuttgart, Germany, 1989.

Beran, R. (1987) "Balanced Simultaneous Confidence Sets," *Biometrika*, **74**, 678-686.

Bhapkar, V. P. (1961) "A Nonparametric Test for the Problem of Several Samples," *The Annals of Mathematical Statistics*, **32**, 1108-1117.

Bhattacharyya, A. (1943) "On a Measure of Divergence Between Two Statistical Populations Defined by Their Probability Distributions," *Bulletin of the Calcutta Mathematical Society*, **35**, 99-109.

Bradley, J. V., *Distribution-Free Statistical Tests*, Prentice-Hall, London, 1968.

Breslow, N. (1970) “A Generalized Kruskal-Wallis Test for Comparing K Samples Subject to Unequal Patterns of Censorship,” *Biometrika*, **57**, 579-594.

Brown, M. B. and Forsythe, A. B. (1974) “Robust Tests for Equality of Variances,” *Journal of the American Statistical Association*, **69**, 364-367.

Brown, G. W. and Mood, A. M. (1951) “On Median Tests for Linear Hypotheses,” *Proc. 2nd Berkeley Symposium Math. Statist. And Prob.*, **2**, 159-166.

Cohen, J., *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed., Erlbaum, Hillsdale, NJ, 1988.

Conover, W. J., Johnson, M. E. and Johnson, M. M. (1981) “A Comparative Study of Tests for Homogeneity of Variances, with Applications to the Outer Continental Shelf Bidding Data,” *Technometrics*, **23**, 351-361.

Cramer, H., *Mathematical Methods of Statistics*, Princeton University Press, 1946.

Davison, A. C. and Hinkley, D. V., *Bootstrap Methods and Their Application*, Cambridge University Press, 1997.

Efron, B. (1979) “Bootstrap Methods: Another Look at the Jackknife,” *The Annals of Statistics*, **7**, 1-26.

Efron, B. and Tibshirani, R. J., *An Introduction to the Bootstrap*, Chapman and Hall/CRC, 1993.

Gehan, E. (1965a) “A Generalized Wilcoxon Test for Comparing Arbitrarily Single Censored Samples,” *Biometrika*, **52**, 203-223.

Gehan, E. (1965b) “A Generalized Two-Sample Wilcoxon Test for Doubly Censored Data,” *Biometrika*, **52**, 650-653.

Graybill, F. A., *Introduction to Linear Statistical Models*, Volume 1, McGraw-Hill, 1961.

Hajek, J., Sidak, Z. and Sen, P. K., *Theory of Rank Tests*, Academic Press, 1999.

Hoeffding, W. (1948) “A class of statistics with asymptotically normal distributions,” *The Annals of Mathematical Statistics*, **19**, 293-325.

Hollander, M. and Wolfe, D. A., *Nonparametric Statistical Methods*, 2nd edition, Wiley-Interscience, 1999.

Hsieh, F. and Turnbull, B., (1996) “Nonparametric and Semiparametric Estimation of the Receiver Operating Characteristic Curve,” *The Annals of Statistics*, **24**, 25-40.

Jonckheere, A. R. (1954) “A Distribution-free K-sample Test Against Ordered Alternatives,” *Biometrika*, **41**, 133-145.

Kang, Q., *Expected Significance Levels and P-Values*, Masters Report, Kansas State University, 2002.

Kruskal, W. H. and Wallis, W. A. (1952) "Use of Ranks in One-Criterion Variance Analysis," *Journal of the American Statistical Association*, **47**, 583-621.

Lawless, J. F., *Statistical Models and Methods for Lifetime Data*, 2nd edition, John Wiley and Sons, 2003.

Lee, A. J., *U-Statistics: Theory and Practice*, New York, Mercel Dekker, 1990.

Levene, H. (1960) "Robust Tests for Equality of Variances," in Ingram Olkin, Harold Hotelling, *et al.*, Stanford University Press, 278-292.

Mandel, M. and Betensky, R. A. (2008) "Simultaneous Confidence Intervals Based on the Percentile Bootstrap Approach," *Computational Statistics and Data Analysis*, **52**, 2158-2165.

Mossman, D. (1999) "Three-way ROCs," *Medical Decision Making*, **19**, 78-89.

Nakas, C. T. and Yiannoutsos, C. T. (2004) "Ordered Multiple-class ROC Analysis with Continuous Measurements," *Statistics in Medicine*, **23**, 3437-3449.

Patil, G. P. and Boswell, M. T. (1970) "A Characteristic Property of the Multivariate Normal Density Function and Some of its Applications," *Annals of Mathematical Statistics*, **41**, 1970-1977.

Peleg, S., Werman, M. and Rom, H. (1989) “A Unified Approach to the Change of Resolution: Space and Gray-level,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **11**, 739-742.

Pepe, M. S. (2000) “Receiver Operating Characteristic Methodology,” *Journal of the American Statistical Association*, **95**, 308-311.

Peto, R. and Peto, J. (1972) “Asymptotically Efficient Rank Invariant Test Procedures,” *Journal of the Royal Statistical Society, Series A (General)*, **135**, 185-207.

Prentice, R. L. (1978) “Linear Rank Tests with Right Censored Data,” *Biometrika*, **65**, 167-179.

Randles, R. H. and Wolfe, D. A., *Introduction to the Theory of Nonparametric Statistics*, New York, Wiley, 1979.

Rizzo, M. L. and Szekely, G. J. (2010) “DISCO Analysis: A Nonparametric Extension of Analysis of Variance,” *Annals of Applied Statistics*, **4**, 1034-1055.

Ross, S. M., *Stochastic Processes*, New York, Wiley, 1996.

Serfling, R. J., *Approximation Theorems of Mathematical Statistics*, Paperback Edition, Wiley-Interscience, 2001.

Snedecor, G. W. and Cochran, W.G., *Statistical Methods*, Eighth Edition, Iowa state University Press, 1983.

Whitney, D. R. (1951) “A Bivariate Extension of the U Statistic,” *The Annals of Mathematical Statistics*, **22**, 274-282.

Xiong, C., VanBelle G., Miller, J. P. and Morris, J. C. (2006) “Measuring and Estimating Diagnostic Accuracy When There are Three Ordinal Diagnostic Groups,” *Statistics in Medicine*, **25**, 1251-1273.