2009-09

Applications of assignment algorithms to nonparametric tests for homogeneity

Ruth, David M.

Monterey, California: Naval Postgraduate School

https://hdl.handle.net/10945/10475

NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

DISSERTATION

APPLICATIONS OF ASSIGNMENT ALGORITHMS TO NONPARAMETRIC TESTS FOR HOMOGENEITY

by

David M. Ruth

September 2009

Dissertation Supervisor: Robert A. Koyak

Approved for public release; distribution is unlimited

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE	*Form Approved OMB No. 0704-0188*

1. AGENCY USE ONLY *(Leave blank)*	**2. REPORT DATE** September 2009	**3. REPORT TYPE AND DATES COVERED** Dissertation

4. TITLE AND SUBTITLE:
Applications of Assignment Algorithms to Nonparametric Tests for Homogeneity

5. FUNDING NUMBERS

6. AUTHOR(S) David M. Ruth

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Naval Postgraduate School
Monterey, CA 93943-5000

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
N/A

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES The views expressed in this dissertation are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (maximum 200 words)

We propose new nonparametric statistical tests to identify whether each element in a sequence of independent multivariate observations is drawn from a common probability distribution or if some distributional change has occurred over the course of the sequence. Each test is formulated using matching techniques based on distances between observations. These tests are capable of detecting changes of quite general nature, and, unlike most similar tests, they require no distribution assumptions or any prior separation of the data into hypothetical pre- and post-change subsets. We derive a central limit theorem for one of the tests and an exact distribution for another. A third culminating test, which is a cumulative sum of statistics on a collection of orthogonal matchings associated with the observation sequence, exhibits noteworthy power to detect whether a distributional change has occurred. We examine the performance of the tests by computer simulation and compare results to a state-of-the-art parametric competitor.

14. SUBJECT TERMS

Nonparametric test, distribution-free test, non-bipartite matching, bipartite matching, change point

15. NUMBER OF PAGES
147

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT Unclassified	**18. SECURITY CLASSIFICATION OF THIS PAGE** Unclassified	**19. SECURITY CLASSIFICATION OF ABSTRACT** Unclassified	**20. LIMITATION OF ABSTRACT** UU

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release; distribution is unlimited

APPLICATIONS OF ASSIGNMENT ALGORITHMS TO NONPARAMETRIC TESTS FOR HOMOGENEITY

David M. Ruth
Commander, United States Navy
B.S. Mathematics, United States Naval Academy, 1991
M.A. Mathematics, University of Texas, 1993

Submitted in partial fulfillment of the
requirements for the degree of

DOCTOR OF PHILOSOPHY IN OPERATIONS RESEARCH

from the

NAVAL POSTGRADUATE SCHOOL
September 2009

Author: ______________________________
David M. Ruth

Approved by:

Robert Koyak
Associate Professor of
Operations Research,
Dissertation Supervisor

Kyle Lin
Associate Professor of
Operations Research

Craig Rasmussen
Associate Professor of
Applied Mathematics

Javier Salmeron
Associate Professor of
Operations Research

Lyn Whitaker
Associate Professor of
Operations Research

Approved by: ______________________________
Robert Dell, Chair, Department of Operations Research

Approved by: ______________________________
Doug Moses, Vice Provost for Academic Affairs

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

We propose new nonparametric statistical tests to identify whether each element in a sequence of independent multivariate observations is drawn from a common probability distribution or if some distributional change has occurred over the course of the sequence. Each test is formulated using matching techniques based on distances between observations. These tests are capable of detecting changes of quite general nature, and, unlike most similar tests, they require no distribution assumptions or any prior separation of the data into hypothetical pre- and post-change subsets. We derive a central limit theorem for one of the tests and an exact distribution for another. A third culminating test, which is a cumulative sum of statistics on a collection of orthogonal matchings associated with the observation sequence, exhibits noteworthy power to detect whether a distributional change has occurred. We examine the performance of the tests by computer simulation and compare results to a state-of-the-art parametric competitor.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

BAP	Bipartite Accumulated Pairs
CUSUM	Cumulative sum
ED	Euclidean distance
ESPM	Ensemble Sum of Pair Maxima
EWMA	Exponentially weighted moving average
GMA	Geometric moving average
JJS	James, James, and Siegmund
MST	Minimum spanning tree
MD	Mahalanobis distance
MD-R	Mahalanobis distance, robust
NAP	Non-Bipartite Accumulated Pairs
NNVE	nearest-neighbor variance estimation
RD	Multivariate rank distance
SPM	Sum of Pair Maxima

THIS PAGE INTENTIONALLY LEFT BLANK

ACKNOWLEDGMENTS

This work is a product of the investments of many. I express my deepest gratitude to those below, without whom this effort would never have succeeded:

I thank God above all for providing me the opportunity to do this work and for granting me all things necessary to bring it to completion.

I thank my wife, Leticia, whose unwavering love, prayer, understanding, and encouragement have been vital to this and any other accomplishment of my adult life. Thanks also to my children, Christian, Eliana, and Elizabeth, for giving my heart cheer during this labor.

I thank my parents, Michael and Barbara Ruth, who have taught me what it means to work hard, equipped me to meet challenges, and supported me in everything I aspire to do.

Thanks to my advisor, Professor Robert Koyak, whose insight, vision, supervision, and guidance have not only made this product what it is, but have shaped me immensely as an academic professional.

I appreciate the contributions of my other committee members, Professors Kyle Lin, Javier Salmeron, Craig Rasmussen, and Lyn Whitaker, who have both improved this work with their input and improved me through their instruction and assistance.

My special thanks go to Professor Samuel Buttrey, who invested substantial time and effort to help me implement state-of-the-art computational routines in order to accomplish the numerical analysis in this study.

I am grateful to Dr. Eric Bechhoefer for his advice and thoughtful contributions to my research and to Goodrich Fuel and Utility Systems for providing funding to support this effort.

I appreciate Dr. Jeffery Haferman and Eric Adint at the Naval Postgraduate School's High Performance Computing Center for their assistance to me as a Hamming-cluster user for my simulation work.

Thanks to the staff at the Dudley Knox Library, who provided me valuable research assistance to complete this work.

Thanks also to Professor Paul Rosenbaum at the Wharton School, University of Pennsylvania, whose contributions in this line of research strongly motivated our work; to Professor David Aldous at the University of California at Berkeley, who suggested Stein's Theorem to us as a possible approach to proving the central limit theorem contained herein; and to Professor Bo Lu at the Ohio State University, who very graciously provided us with efficient computational routines to implement some of our ideas.

Finally, I am thankful for the support of my office-mates over the last three years: Hiro Sato, Ali Al-Rowaei, Eng Yau Pee, Jay Foraker, and Anthony Tvaryanas. They not only sharpened my thinking, but also made this labor more fulfilling by their friendship.

EXECUTIVE SUMMARY

Given a sequence of observations, has a change occurred in the underlying probability distribution with respect to observation order? This question arises in a variety of applications, including quality control, machinery health diagnosis, biosurveillance, and image analysis. This problem is encountered throughout statistical literature and is often referred to as "the change-point problem," where "change point" refers to the index of the first observation for which the underlying probability distribution is different from that of previous observations. Detecting change points in high-dimensional settings is challenging, and most change-point methods for multi-dimensional problems rely heavily upon distributional assumptions such as multivariate normality or the use of observation history to model probability distributions. In practice, such strong distributional assumptions are often invalid and can lead to poor detection performance, and useful observation histories are often unavailable. Also, most change-point methods are applicable only to changes of a specific type (for example, an abrupt change in distribution mean) when in many cases one is interested in detecting more general types of change as well (for example, changes in scale or gradual changes).

We propose new nonparametric statistical tests to detect the presence of a change point in a sequence of multivariate data based on the graph-theoretic concept of matching. Each test requires only the assumption of some reasonable function to measure dissimilarity between observations. We state the change-point problem by representing the observation set as a complete graph in which each observation is a vertex and each pair of vertices is connected by an edge whose weight is the dissimilarity between the two vertices. Then we pair observations together in such a way as to minimize the total cost of the collection of pairs; this collection of pairs is called a matching. Our statistical tests for a change point use the fact that if a change has occurred with respect to order in the underlying distribution of a sequence of observations then the sequence labels of the pairs in the matching are closer together on average than if no distributional change has occurred. By considering not just the lowest-cost matching but rather several low-cost matchings, we achieve considerable power to detect whether there is a change point in a sequence of observations.

We examine the performance of these tests by simulation in various change-point settings considering different underlying probability distribution family, dimensionality, change point location, change parameter (distribution location or scale), type of change (abrupt or gradual), and magnitude of change. Each test demonstrates power to detect a change point at fixed false alarm rate in every case examined. The most powerful of these tests is the Ensemble Sum of Pair-Maxima (ESPM) test, which computes the cumulative sum of the larger sequence labels among all pairs in a collection of low-cost matchings and measures the deviation of this sum from its expected value. The ESPM test has change detection power comparable to a state-of-the-art parametric competitor, the maximum likelihood ratio test of James, James, and Siegmund (JJS), even when the parametric assumptions for that test are met. When those assumptions are not met, the ESPM test retains noteworthy power to detect a change point, while the false alarm rate of the JJS test increases.

I. INTRODUCTION

The problem of identifying a change in a stochastic process is often referred to as "the change-point problem" and has been a subject of enduring interest in statistical literature. A simple statement of this problem is, "Given a sequence of observations, has a change occurred in the underlying probability distribution with respect to observation order?" This problem arises in a variety of applications, such as:

- *Quality control*. Samples are taken from a particular manufacturing process, perhaps over time or across different stages in the process, that carry information regarding the quality of the end product. Change-point methods are used to indicate if, where, or when the process departs from an "in-control" condition.

- *Machinery diagnostics and fault detection*. Consider a complex machine for which various measurements are taken during its operation that provide an indication of machine health. A change in the distribution of these measurements with respect to time might indicate some form of health degradation.

- *Biosurveillance*. Suppose occurrences of a particular disease are cataloged by geographic location and monitored over time. Change point methods may be used to detect whether a disease outbreak has occurred.

- *Image analysis*. Consider a sequence of images of the same scene taken at different times: for example, satellite images of some geographic area or magnetic resonance imaging scans on an individual. Evidence that the scene is changing in some significant way may be found by means of change point analysis.

This research is about detecting whether change has occurred. Specifically, we investigate *nonparametric* tests to detect change points of *very general nature* in *multivariate* data. Cases of "very general nature" include those of abrupt or gradual changes in the mean of a distribution, or changes in its scale. While many real-life processes exhibit gradual change, fairly little investigation has been done in the area of detecting gradual changes in multivariate data. Furthermore, a relatively small body of work exists proposing nonparametric solutions to multivariate change detection problems, although interest in this area appears to be growing.

We examine nonparametric methods that rely on *matching*, which involves the pairing together of observations based on some measure of dissimilarity. Existing statistical applications of matching techniques include:

- assessing sensor accuracy in test and evaluation of radar and joint-tracking systems by pairing detected objects with truth objects,
- comparing the accuracy of different methods for estimating the locations of impact points in munitions testing by pairing estimated locations to "ground truth," and
- pairing subjects based on similarity measures for clinical trials and observational studies,

to name a few. In this paper, we introduce new methods of this type that prove to be powerful to detect change over a wide range of alternative hypotheses.

Our work is organized as follows: In Chapter II we classify the field of change-point problems into its two main categories, sequential and non-sequential techniques, with a formal discussion of how problems in each category are framed. We then summarize our review of the literature in this field, with particular emphasis on nonparametric approaches, followed by a graph-theoretic overview of matching. The chapter concludes with a review of the most recent work on this problem based on non-bipartite matching (defined in the next chapter), which is our primary area of interest here. In Chapter III, we propose new statistical tests based on non-bipartite matching. First, we introduce the Sum of Pair-Maxima (SPM) test and the Non-Bipartite Accumulated Pairs (NAP) test, and develop the supporting theory for these tests in some detail. Of primary importance are the proof of a central limit theorem for the SPM test and the derivation of the exact distribution for the NAP test. Then we introduce the dominant test of this paper, the Ensemble Sum of Pair-Maxima (ESPM) test, which is an extension of the SPM test that involves extracting additional change-point information from orthogonal matchings. Finally, we present the Bipartite Accumulated Pairs (BAP) test as an application of bipartite matching techniques to change-point problems where some observation history is available. Chapter IV demonstrates the performance of the SPM, NAP, and ESPM tests by means of a simulation study. The power of these tests is compared to a state-of-the-art parametric competitor, the maximum likelihood ratio test of James, James, and Siegmund (1992), for various cases including different underlying distributions, different types and magnitudes of change, and different dissimilarity measures. Chapter V summarizes our findings and outlines opportunities for further research in this field.

THIS PAGE INTENTIONALLY LEFT BLANK

II. PROBLEM BACKGROUND AND LITERATURE REVIEW

A. PROBLEM FORMULATION

This research addresses the following specific question: Given a sequence of independent multivariate observations, is the sample statistically homogeneous? In other words, has the underlying probability distribution from which the observations were drawn *remained constant* or has it *changed*? As discussed in the previous chapter, this problem emerges in a wide variety of applications and is traditionally referred to as "the change-point problem."

1. Change Points

We define the term **change point** as follows: Given a sequence of independent random variables $\left(X_1, X_2, \ldots, X_N\right)$, let F_i denote the probability distribution of X_i. Then an integer $\tau \in \{2, \ldots, N\}$ is a change point with respect to measure δ if $F_1 = F_2 = \cdots = F_{\tau-1}$, $F_{\tau-1} \neq F_\tau$ and $\delta\left(F_1, F_j\right) - \max_{k \in \{\tau, \ldots, j\}} \delta\left(F_k, F_j\right)$ is strictly positive over $j \in \{\tau, \ldots, N\}$, where δ is a measure of distance between two probability distributions. The F_j ($j \geq \tau$) are not necessarily distinct from one another; for example, the distribution change at τ may be associated with an abrupt mean change, or "mean jump." Or F_j may some simple function of $j \geq \tau$; for example, the distribution change beginning at τ may be associated with a gradual mean change, or "mean drift." More complicated forms for F_j are allowed as well. A formulation of the general change-point problem in a hypothesis-testing framework with respect to observations $X_1, X_2, \ldots, X_N$ consists of defining null hypothesis

$$H_0: \quad F_1 = F_2 = \cdots = F_N \tag{2.1}$$

and corresponding alternative hypothesis

H_1 : There exists an integer τ, $2 \le \tau_0 \le \tau \le \tau_1 \le N$, such that

(2.2) $F_1 = F_2 = \cdots = F_{\tau-1}$, $F_{\tau-1} \ne F_\tau$, and

$\delta\left(F_1, F_j\right) - \max_{k \in \{\tau,\ldots,j\}} \delta\left(F_k, F_j\right)$ is strictly positive over $j \in \{\tau, \ldots, N\}$.

Usually we take $\tau_0 = 2$ and $\tau_1 = N$ but in some cases may wish to restrict the change point to a narrower interval.

An important taxonomy exists within the family of change-point problems; we present a particular classification scheme here, similar to Brodsky and Darkhovsky (1993) and Basseville and Nikiforov (1993), to serve as a concept map of sorts, to provide a framework for review of the literature in this field, and to make clear the classification of the problem for which we offer solutions.

2. Taxonomy of Change-Point Problems

Change-point problems can be classified into the two broad categories of *sequential* and *non-sequential analysis*. Sequential analysis involves detecting the occurrence of a change while monitoring a system “on-line;” that is, data collection is ongoing in time and analysis is performed sequentially as the data set is updated. For such cases, the null hypothesis is tested over and over again as new data are added to the set of observations as follows:

1) With observations $X_1, \ldots, X_{t-1}$ on hand, add X_t.

2) Test for a change point in $\{2, \ldots, t\}$.

    a) If a change point is detected, perform some predetermined action.

    b) If not, go back to step (1) with $X_1, \ldots, X_t$ on hand.

A classic application of sequential analysis is in the arena of quality control, where some particular process is ongoing in time and process outputs are sampled and tested in sequence to identify if some undesirable change in the process has occurred.

In contrast, non-sequential analysis involves the examination of a finite sequence of data "off-line" with the purpose of identifying whether or not some change has occurred during the observation period. Non-sequential analysis can be divided further into the cases of a *single test* or *simultaneous tests.* For a single test, one considers a sequence of multivariate observations $(X_1, X_2, \ldots, X_N)$ and a known or assumed time $\tau \in \{2, \ldots, N\}$ and then tests the null hypothesis $H_0: F_1 = F_2 = \cdots = F_N$ against the alternative $H_1: F_1 = F_2 = \cdots = F_{\tau-1} \neq F_\tau = F_{\tau+1} = \cdots = F_N$ (or perhaps a one-sided alternative). One example of this type of problem is the clinical trial scenario, where two groups of subjects are drawn from some common population, one group is administered a treatment and the other a placebo, and the problem is to test whether the treatment has some particular effect. In the single test framework, *N* specimens are split into a control and a test group with $\{X_1, \ldots, X_{\tau-1}\}$ and $\{X_\tau, \ldots, X_N\}$ being the two associated sets of observations (the order of observations within groups does not matter for this case), and a single hypothesis test is performed regarding τ (hence the name "single test").

For simultaneous tests, no specific candidate for change point τ is assumed; the null hypothesis (2.1) is tested against alternative (2.2) as stated. Such tests are simultaneous in the sense that they involve testing all possible change points simultaneously. An example of such a test relates to machinery health management: Consider a military aircraft with an on-board device that records various measurements on the aircraft with respect to time during flight; these measurements are known or believed to be indicators of aircraft health. Assume for the sake of this example that the observations are independent with respect to time. Let $(X_1, X_2, \ldots, X_N)$ be the sequence of these observations. When the aircraft returns from its mission, these observations are analyzed for evidence of health degradation. That is, do the observations provide evidence of a change from some "healthy" distribution to a "less healthy" one? If so, may one infer *when* the degradation occurred (or began to occur)?

Machinery health diagnosis and prognosis problems are strong motivations for our research effort; consequently, the focus of this research is non-sequential simultaneous testing. Our literature survey finds that powerful *nonparametric* tests for *multivariate* change-point problems are not abundantly available, particularly in the non-sequential simultaneous testing case. We now proceed to review several parametric and nonparametric approaches to univariate and multivariate change-point problems, concluding with a discussion of the graph-theoretic concept of *matching* and its application to such problems. This will set the stage for our introduction in the next chapter of new matching-based solutions to the change-point problem.

B. PARAMETRIC APPROACHES

1. Univariate Case

The two-sample *t*-test, described in Tanis and Hogg (2008), is perhaps the most widely known test for heterogeneity, though it generally is not presented in change-point terms. It is one of the first tests introduced in an undergraduate statistics course, and it usually is framed as a test for a difference in the mean of two samples. While it is widely applicable, it has the obvious limitations that 1) it applies only the univariate case, 2) it assumes the underlying distributions are normal, and 3) it only tests for differences in distribution means.

In quality assurance, the classic univariate sequential test for a change point is the cumulative sum (CUSUM) test introduced by Page (1954). Others include the sequential *t*-test (Rushton, 1950), Geometric Moving Average (GMA) or Exponentially Weighted Moving Average (EWMA) procedures (Roberts, 1959), and Bayes-type procedures (Girshick and Rubin, 1952; Shirayev, 1963). These procedures are powerful in many settings and can be applied as non-sequential or sequential change-point tests. Of course these tests are also limited to univariate cases, although some multivariate extensions exist. Additionally, they require assumptions about the underlying data distribution (normality is usually assumed).

2. Multivariate Case

The multivariate analog of the two-sample *t*-test is Hotelling's two-sample T^2 test (Hotelling, 1931), which detects differences in the mean vectors of two multivariate normal samples. Hotelling's T^2 statistic is a non-sequential single test. Sequential multivariate parametric tests include multivariate extensions of CUSUM procedures (Basseville and Nikiforov, 1993; Runger and Testik, 2004) and EWMA procedures (Lowry *et al.*, 1992; Prabhu and Runger, 1997).

A change-point test by James, James, and Siegmund (hereafter, "JJS") (1992) associated with an abrupt change in the mean of a multivariate normal distribution interests us particularly. JJS is a non-sequential simultaneous test—the area of our interest in this work—and it serves as a powerful competitor to methods we present later. Given multivariate observations $X_1, X_2, \ldots, X_N$, JJS uses the modified likelihood ratio test statistic

$$(2.3) \qquad T_{\text{JJS}} = \max_{k_0 \le k \le k_1} \frac{N}{k(N-k)} \left(S_k - \frac{k}{N} S_N \right)' \left(U_N - \frac{1}{N} S_N S_N' \right)^{-1} \left(S_k - \frac{k}{N} S_N \right)$$

where $S_k = \sum_{i=1}^{k} X_i$, $U_k = \sum_{i=1}^{k} X_i X_i'$, and k_0 and k_1 are the lower and upper limits, respectively, of the interval possibly containing a change point. The actual likelihood test is based on the case $k_0 = 1$ and $k_1 = N - 1$ (that is, the change point could be anywhere in the observation sequence), but the test provides the flexibility to limit the search for a change point to a subinterval of $(1, N-1)$. JJS show that the tail probabilities for this test can be well-approximated by

$$(2.4) \qquad \begin{aligned} & P(T_{\text{JJS}} \ge x) \cong \\ & \frac{1}{\Gamma(p/2)} \left(\frac{Nx}{2} \right)^{p/2} (1-x)^{(N-p-3)/2} \int_{t_0}^{t_1} \frac{1}{t(1-t)} \nu \left(\left(\frac{x}{t(1-t)(1-x)} \right)^{1/2} \right) dt, \end{aligned}$$

where $k_0 / N \to t_0$ and $k_1 / N \to t_1$ as $N \to \infty$, $0 < t_0 \le t_1 < 1$, and

$$(2.5) \qquad \nu(t) = \frac{2}{t^2} \exp\left\{ -2 \sum_{k=1}^{\infty} \frac{1}{k} \Phi\left(-\frac{t k^{1/2}}{2} \right) \right\},$$

with Φ being the standard normal cumulative distribution function. The integral term in (2.4) is computed satisfactorily by numerical methods. JJS present a heuristic modification to improve their tail probability approximation; we do not describe those details here, but we do utilize the modification of the JJS test for comparison purposes later.

A drawback to all these parametric approaches to change-point detection is that they are not necessarily robust across different underlying data distributions. In particular, the assumption of multivariate normality in many cases proves to be difficult to justify. We proceed to review some existing nonparametric approaches to the change-point problem.

C. NONPARAMETRIC APPROACHES

1. Univariate Case

A classic nonparametric test to determine whether two univariate random samples come from the same population is the Mann-Whitney test (Mann and Whitney, 1947), also known as the Wilcoxon rank sum test (Conover, 1999). Let $\mathsf{A}_1 = \left\{X_1, \ldots, X_m\right\}$ and $\mathsf{A}_2 = \left\{X_{m+1}, \ldots, X_{m+n}\right\}$ be two sets of observations with all X_i being members of some ordered set. Assign ranks (or midranks in the case of ties) to the observations with respect to set ordering and let $R\left(X_i\right)$ denote the rank of observation X_i. Intuitively, one would expect that if the observations in A_1 tend to be smaller than those in A_2 then the ranks of the observations in A_1 would be smaller on average than the ranks of those in A_2. To determine whether A_1 and A_2 are drawn from the same population one computes the Mann-Whitney test statistic, which is simply

$$(2.6) \qquad T_{\mathrm{MW}} = \sum_{i=1}^{m} R\left(X_i\right).$$

Let F_1 and F_2 be the distribution functions corresponding to the observations in A_1 and A_2, respectively, and let $Y \sim F_1$ and $Z \sim F_2$. Then the hypotheses associated with the Mann-Whitney test may be stated as follows.

$$\begin{aligned} &H_0: \quad P(Y > Z) = 0.5; \\ &H_1: \quad P(Y > Z) \neq 0.5. \end{aligned} \tag{2.7}$$

For small samples, quantiles of T_{MW} are found in tables or by using standard functions in statistical software (such the "qwilcox" function in R (2005)). For large samples T_{MW} is asymptotically normal. While the Mann-Whitney test is consistent against mean difference alternatives, it is not sensitive to other types of differences (for example, differences in scale).

Another rank-based non-sequential single test, which is consistent against a broader range of alternatives but is less powerful than the Mann-Whitney test (Conover, 1999), is the Wald-Wolfowitz runs test (Wald and Wolfowitz, 1940). Observation ranks are computed as above, but this time the ranks are collected into runs of consecutive ranks that come from the same group (either A_1 or A_2). The test statistic is simply the number of runs in the collection; a relatively small number of runs indicates that the two samples are from different distributions. Like the Mann-Whitney test, the Wald-Wolfowitz test is asymptotically normal.

Two other tests that are consistent against any type of difference that might exist between underlying distributions are the Kolmogorov-Smirnov test and the Cramér-von Mises test. If S_1 and S_2 are the empirical distribution functions for the observations in A_1 and A_2, respectively, then the Kolmogorov-Smirnov test statistic is given by

$$T_{\mathrm{KS}} = \sup_x \left| S_1(x) - S_2(x) \right| \tag{2.8}$$

and the Cramér-von Mises test statistic is given by

$$T_{\mathrm{CvM}} = \frac{mn}{(m+n)^2} \sum_{x \in \mathsf{A}_1 \cup \mathsf{A}_2} \left[S_1(x) - S_2(x) \right]^2 . \tag{2.9}$$

In other words, these tests statistics evaluate the supremum norm and L^2 norm, respectively, of the difference between two empirical distribution functions. Large values of these statistics are evidence that A_1 and A_2 are drawn from different probability distributions.

These four particular tests broadly represent the two primary techniques used for nonparametric change-point detection both in univariate and multivariate cases: techniques based on rank permutations (such as Mann-Whitney and Wald-Wolfowitz) and tests based on distribution function estimation (such as Kolmogorov-Smirnov and Cramér-von Mises). The new methods we present in the next chapter are grounded in rank permutation arguments.

Nonparametric sequential tests include generalizations of CUSUM procedures such as those introduced by Bhattacharya and Frierson (1981) and Gordon and Pollock (1995). These procedures apply to the univariate case only; no extensions of these tests to the multivariate case have yet been proposed (Fricker and Chang, 2009).

2. Multivariate Case

a. General Approaches

Multivariate extensions do exist for both the Kolmogorov-Smirnov and Cramér-von Mises tests. For example, Bickel (1969) extends the Kolmogorov-Smirnov test to $\mathbb{R}^d$ by defining multivariate rank vectors, computing empirical distribution functions with respect to within-group multivariate ranks, and then evaluating the supremum norm on the difference of the two empirical distribution functions as a test statistic. Præstgaard (1995) extends Bickel's result to more general sample spaces. Baringhaus and Franz (2004) propose a Cramér-von Mises-like statistic on $\mathbb{R}^d$ by comparing the average Euclidean distance between points in different groups to the average distance between points in the same group. Hall and Tajvidi (2002) propose a permutation test using a nearest-neighbors approach that generalizes both the Kolmogorov-Smirnov and Cramér-von Mises tests to the multivariate case. These tests all rely on simulation to compute estimated quantiles for the test statistic null distribution.

Li and Liu (2004) apply the notion of data depth to nonparametric tests for changes in multivariate location or scale. Data depth is a way of measuring how "deep" or "central" a point is with respect to a particular distribution or sample (Liu *et al.,* 1999). Well-known examples of such measures in the data depth literature include half-space depth (Hodges (1955) , Tukey (1975), sometimes referred to as Tukey depth), convex hull peeling depth (Barnett, 1976), and simplicial depth (Liu, 1990). Data depths provide a natural center-outward ordering of points in a multivariate sample; once ordered, various univariate tests for change may be applied with respect to the ordering.

Fricker and Chang (2009) propose a sequential change-point test which uses an available history of multivariate observations combined with the *k* most recent observations (where *k* is an adjustable window parameter) to generate a nonparametric running estimate of the underlying density distribution. The history is assumed to be in control; that is, the historical observations are all drawn from the null distribution with no change point. With each new observation they compute a new density estimate and then perform a Kolmogorov-Smirnov test to identify whether the density heights of the data of interest are uniformly distributed among the density heights of the historical data.

b. Graph-Theoretic Approaches and Matching

An intriguing approach to the change-point problem involves applying graph-theoretic ideas. In particular, methods based on the graph-theoretic concept of matching have gained interest in recent years, in no small part due to increases in computational capacity. The test statistics we propose in this work are all matching-based; therefore, we conclude this chapter by providing necessary graph theory definitions and background to develop our ideas and reviewing graph-theoretic approaches introduced by Friedman and Rafsky (1979) and Rosenbaum (2005) which have inspired our work.

(1) Definitions. The definitions in this section are from Chartrand and Zhang (2005). A **graph** *G* consists of a finite nonempty set *V* of elements called **vertices** and a set *E* of two-element (unordered) subsets of *V* called **edges**, in which case we write $G=(V,E)$. A graph $G_1=(V_1,E_1)$ is called a **subgraph** of $G=(V,E)$ if $V_1 \subseteq V$ and $E_1 \subseteq E$; if $V_1 = V$ then G_1 is called a **spanning subgraph** of G. We denote the edge joining vertices u and v by $\{u,v\}$. Two distinct vertices are **adjacent vertices** if they are joined by an edge, and two distinct edges are **adjacent edges** if they share a

vertex. Vertex u and edge $\{u,v\}$ are said to be **incident** with each other, and the **degree** of vertex u is the number of edges incident with u.

A $u-v$ **walk** in graph G is a sequence of vertices in G beginning with u and ending with v such that consecutive vertices in the sequence are adjacent; if $u=v$ then the walk is **closed.** A $u-v$ **trail** is a walk in which no edge is traversed more than once; a **circuit** is a closed trail including at least three distinct vertices. A circuit that repeats no vertex except for the first and last is a **cycle**. If there exists a $u-v$ **walk** for every pair of vertices u and v in graph G then G is said to be **connected**.

A graph G is called **acyclic** if it has no cycles, a **tree** is an acyclic connected graph, and a **spanning tree** of G is a spanning subgraph of G that is a tree. If a real number is assigned to each edge of a graph, then the graph is a **weighted graph** and the sum of the all the edge weights is called the **weight** of the graph. A spanning tree of weighted graph G whose weight is smallest among all spanning trees of G is called a **minimum spanning tree** (MST).

Friedman and Rafsky (1979) consider various statistics based upon MSTs in order to test whether two samples are drawn from the same distribution. Given sets of observations A_1 and A_2 as above, they construct a MST with respect to some cost function on the sample space. One change-point detection method they consider consists of removing each edge of the MST that connects a point in A_1 to a point in A_2, and then defining a test statistic that counts the number of disjoint subtrees that result from the edge removal. The resulting test is effectively a multivariate runs test, which corresponds exactly to the Wald-Wolfowitz runs test for the univariate case. The Wald-Wolfowitz runs test is known to be not particularly powerful (Connover, 1999, p. 3). Friedman and Rafsky (1999) demonstrate that their multivariate runs test has high power in higher dimensions, and they enhance test power by computing their test statistic on a collection of **orthogonal** MSTs, where two MSTs are orthogonal if they have no edges in common. We will use a similar idea to extend our main results for improved power.

We require a few final definitions to develop our main results. A subset of edges $E' \subseteq E$ is **independent** if no two edges in E' are adjacent. A **matching** in a graph $G = (V, E)$ is an independent set of edges in *G*. A **maximum matching** in *G* is a matching that consists of at least as many edges as any other possible matching in *G*. For the remainder of this paper, all matchings under consideration are maximum matchings (that is, we are interested only in matchings that pair together as many vertices as possible). Finally, a **perfect matching** in *G* is a matching that includes every vertex of *G*. A perfect matching is necessarily a maximum matching; furthermore, a perfect matching is possible only on graphs with an even number of vertices.

(2) Bipartite and Non-Bipartite Matching. A variety of problems can be framed as matching problems, where a matching is sought that minimizes some cost (Ahuja *et al.,* 1993, p. 9). Two specific cases are **bipartite matching** problems, where graph vertices are divided into two distinct subsets $\mathcal{A}_1$ and $\mathcal{A}_2$ and each edge consists of one vertex each from $\mathcal{A}_1$ and $\mathcal{A}_2$, and **non-bipartite matching** problems, where the matching does not depend on a partition of the vertices (that is, any two vertices may be paired in the matching).

In our case, we are interested in matchings of minimum **cost**, where a cost function is defined as follows: Given sample space $\mathcal{S}$, $c : \mathcal{S} \times \mathcal{S} \to [0, \infty)$ is a cost function if it satisfies

$$c(x,x) = 0 \ \ \forall x \in \mathcal{S} \tag{2.10}$$

and

$$c(x,y) = c(y,x) \ \forall x, y \in \mathcal{S} \ , \tag{2.11}$$

We use c_{ij} to denote the cost $c(X_i, X_j)$ and also sometimes use the common terminology that c_{ij} is the weight of the edge joining X_i and X_j. We will call cost function c a **distance** function if it satisfies the triangle inequality

$$c(x,z) \leq c(x,y) + c(y,z) \ \forall x, y, z \in \mathcal{S} \tag{2.12}$$

in addition to (2.10) and (2.11) . In general, this framework allows broad flexibility to accommodate all types of data (discrete or continuous, numeric or categorical, etc.). In the change-point setting, a cost function should be tailored in some sensible way to the data types of interest and its selection ultimately does matter in detecting departures from the null hypothesis.

We formulate the minimum-weight *bipartite* matching problem in the following manner. Given sample space $\mathcal{S}$, two distinct sets of observations $\mathsf{A}_1 = \{X_1, \ldots, X_m\} \subseteq \mathcal{S}$ and $\mathsf{A}_2 = \{X_{m+1}, \ldots, X_{m+n}\} \subseteq \mathcal{S}$, $N = m + n$, and cost function c, a minimum-weight bipartite matching is a solution to the problem

$$\begin{array}{lll} & \underset{x}{\text{minimize}} & \sum_{i=1}^{m} \sum_{j=m+1}^{N} c_{ij} x_{ij} \\ & \text{subject to} & \sum_{j=m+1}^{N} x_{ij} \le 1 \quad \forall i \in \{1, \ldots, m\}, \\ (2.13) & & \sum_{i=1}^{m} x_{ij} \le 1 \quad \forall j \in \{m+1, \ldots, N\}, \\ & & \sum_{i=1}^{m} \sum_{j=m+1}^{N} x_{ij} = \min(m, n), \\ & & x_{ij} \in \{0,1\} \quad \forall i \in \{1, \ldots, m\}, \forall j \in \{m+1, \ldots, N\}, \end{array}$$

where $x_{ij} = 1$ indicates that edge $\{X_i, X_j\}$ is in the matching and $x_{ij} = 0$ otherwise. In the graph underlying this problem, each element of A_1 is joined by an edge to each element of A_2; no edges join elements within A_1 or within A_2. The solution to this problem is not necessarily unique. In operations research literature, this problem is a particular instance of the "general assignment problem" (Ahuja *et al.,* 1993, pp. 639-640). Software algorithms to solve this problem include that of Jonker and Volgenant (1987) and are widely available.

Alternatively, we formulate the minimum-weight *non-bipartite* matching problem as follows. Given sample space $\mathcal{S}$, a single set of observations $\mathcal{A} = \{X_1, \ldots, X_N\} \subseteq \mathcal{S}$, and cost function c, a minimum weight non-bipartite matching is a solution to the problem

$$
\text{(2.14)} \qquad
\begin{aligned}
\underset{x}{\text{minimize}} \quad & \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} c_{ij} x_{ij} \\
\text{subject to} \quad & \sum_{i=1}^{k-1} x_{ik} + \sum_{j=k+1}^{N} x_{kj} \le 1 \quad \forall k \in \{1, \ldots, N-1\}, \\
& \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} x_{ij} = \lfloor N/2 \rfloor, \\
& x_{ij} \in \{0,1\} \quad \forall j \in \{i+1, \ldots, N\}, \forall i \in \{1, \ldots, N-1\}.
\end{aligned}
$$

where $\lfloor y \rfloor$ is the smallest integer less than or equal to y and x_{ij} indicates whether $\{X_i, X_j\}$, $i < j$, is in the matching as in the bipartite matching case. In this case, every pair of elements in A is joined by an edge and the underlying graph is referred to as a **complete graph.** The solution to this problem is a matching that consists of $N/2$ edges with every vertex included in the matching if N is even, or $(N-1)/2$ edges with every vertex but one included in the matching if N is odd. As in the bipartite matching case, the solution to the non-bipartite matching problem is not necessarily unique. Algorithms by Edmonds (1965) and Derigs (1988) solve the minimum-weight non-bipartite matching problem on a complete graph in $O(N^3)$ time. Several improvements to Edmond's algorithm have been developed over the years (Gabow, 1973; Galil *et al.*, 1986; Gabow *et al.,* 1989; Cook and Rohe, 1999; Mehlhorn and Schäfer, 2002; Kolmogorov, 2009). While these improvements do not improve theoretical runtime on a complete graph, they have been shown to improve realized runtimes in many practical instances. Edmond's original algorithm to solve the non-bipartite weighted matching problem is sometimes called Edmond's "blossom" algorithm, where the term "blossom" refers to subgraphs with particular properties that emerge during execution of the algorithm. Subsequent improvements often carry the "blossom" moniker; the latest is Kolmogorov's "Blossom V" (2009). In Chapter IV, we elaborate on computational details specific to our implementation of non-bipartite matching algorithms for this study.

(3) Cost Functions. In both the bipartite and non-bipartite cases, the assignment of pairs in a matching depends upon the choice of a cost function. For the problem of testing for data homogeneity with respect to some ordering, we will use cost

functions that are reasonable dissimilarity measures. Some applications will invite a natural choice of dissimilarity measure; for other applications this choice may require some deliberation. In our simulation study we consider four different distance functions on the sample space $S = \mathbb{R}^d$. The first is Euclidean distance (ED), which yields

$$c_{ij}^{\mathrm{ED}} = \left[\left(X_i - X_j \right)' \left(X_i - X_j \right) \right]^{1/2} . \tag{2.15}$$

One disadvantage of ED is that it does not take into account measurement scale or correlation among data components. To address these issues, Mahalanobis distance (MD) is often used as an alternative, which is scale-invariant and accounts for data correlation. The resulting cost function is

$$c_{ij}^{\mathrm{MD}} = \left[\left(X_i - X_j \right)' V^{-1} \left(X_i - X_j \right) \right]^{1/2} , \tag{2.16}$$

where V is an estimate of the covariance matrix associated with the data of interest. Estimating V by the sample covariance matrix is sensitive to outliers, however, so we are also interested in distance measures that are robust to such outliers. Wang and Raferty (2002) provide a useful method to downweight outliers, called nearest-neighbor variance estimation (NNVE) that measures how outlying a data point is by the standardized distance between the point and its k^{th} nearest neighbor (where k is an adjustable parameter). Therefore, we consider a third distance function, which we will refer to as "Mahalanobis distance, robust" (MD-R), which is simply the MD computed with reference to the NNVE covariance estimate V_{NNVE} (we omit any k subscript to simplify notation here):

$$c_{ij}^{\mathrm{MD}} = \left[\left(X_i - X_j \right)' V_{\mathrm{NNVE}}^{-1} \left(X_i - X_j \right) \right]^{1/2} . \tag{2.17}$$

Another way to reduce sensitivity to outliers is to consider a distance measure based on the idea of multivariate ranks. We adopt a useful extension of rank to higher dimension given by Barnett (1976) and Chaudhuri (1996), described as follows by Choi and Marden (1997). Consider a d-dimensional inner product space S

and observations $X_1,\ldots,X_N \in \mathcal{S}$. For $d=1$, denote the rank of observation X_i by $r^{(i)}$, assigning midranks in the case of ties. Center and scale these ranks by the transformation

$$R(X_i)=\frac{2r^{(i)}-n-1}{n}, \tag{2.18}$$

which maps $\left\{r^{(1)},\ldots,r^{(N)}\right\}$ into the open interval $(-1,1)$. Then (2.18) can be written

$$R(X_i)=\frac{1}{n}\sum_{j=1}^{n}\operatorname{sgn}\left(X_i-X_j\right), \tag{2.19}$$

where sgn is the signum function

$$\operatorname{sgn}(x)=\begin{cases}0 & \text{if } x=0;\\ \dfrac{x}{|x|} & \text{otherwise.}\end{cases} \tag{2.20}$$

Now for $X_i \neq X_j$, the summand in (2.19) can be expressed as

$$\operatorname{sgn}\left(X_i-X_j\right)=\frac{X_i-X_j}{\left|X_i-X_j\right|}. \tag{2.21}$$

Then a very natural extension of this centered and scaled ranking transformation to the case $d>1$ is obtained by defining

$$R(X_i)=\frac{1}{n}\sum_{j\neq i}\frac{X_i-X_j}{\left\|X_i-X_j\right\|}, \tag{2.22}$$

where $\|\cdot\|$ may be any norm in general; unless otherwise specified we will use $\|X\|=\sqrt{X'X} \quad \forall X \in \mathcal{S}.$ We will use the term “multivariate rank distance” (RD) to refer to the Euclidean distance between multivariate ranks as our fourth and final distance option:

$$c_{ij}^{\mathrm{RD}}=\left[\left(R(X_i)-R(X_j)\right)'\left(R(X_i)-R(X_j)\right)\right]^{1/2}. \tag{2.23}$$

In Chapter IV, we examine the impact these different distance functions have on our ability to detect change with the new change-point detection methods presented in the next chapter.

(4) Matching Examples. We now illustrate these matching concepts with a few simple examples. Figures 1, 2, and 3 show the same 20 observations, each drawn from a bivariate normal distribution, with Euclidean distance as the cost function. The coordinates for these data are listed in Appendix D. In Figure 1, the observations are randomly partitioned into two subsets containing 8 and 12 observations indicated by group labels '□' and '*', respectively; the resulting minimum-weight bipartite matching consists of 8 pairs, with 4 observations from the larger set remaining unmatched. A quite different matching results when the observations are partitioned in a different way, as shown in Figure 2 with two subsets of equal size. Finally, Figure 3 demonstrates the minimum-weight non-bipartite matching on these same 20 points with no prior partitioning. These examples reiterate two fundamental differences between bipartite and non-bipartite matchings. First, the non-bipartite matching is not associated with any prior partitioning of the observation set, while the bipartite matching depends greatly on how the observation set is partitioned. Indeed, for an even number of observations one may think of the minimum-weight non-bipartite matching as the lowest cost matching among all minimum-weight *bipartite* matchings computed for all possible partitions of equal size. Second, in the case of non-bipartite matching for an even number of observations, *all* observations are paired with another (all but one are paired if the number of observations is odd), while in the bipartite matching case some observations are necessarily left unmatched in the larger of the partition sets.

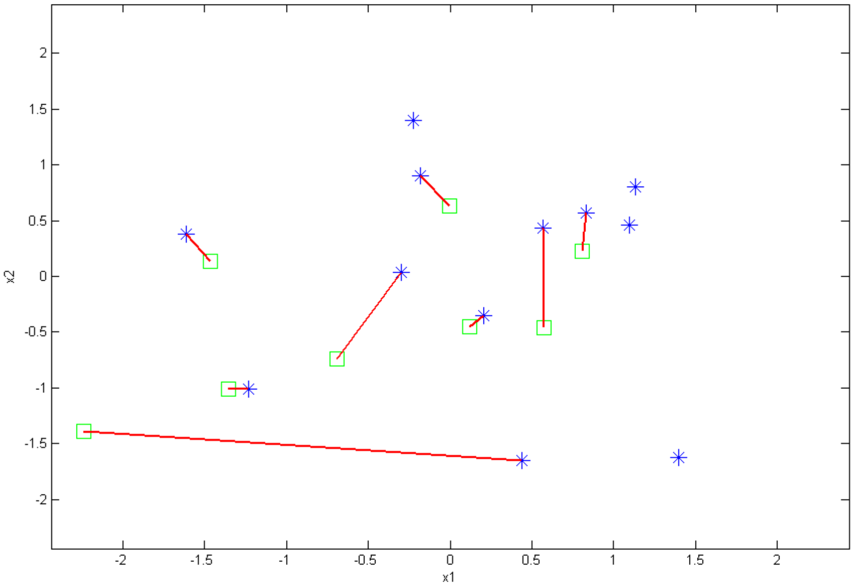

Figure 1. Minimum-weight bipartite matching on 20 points; *m*=8, *n*=12.

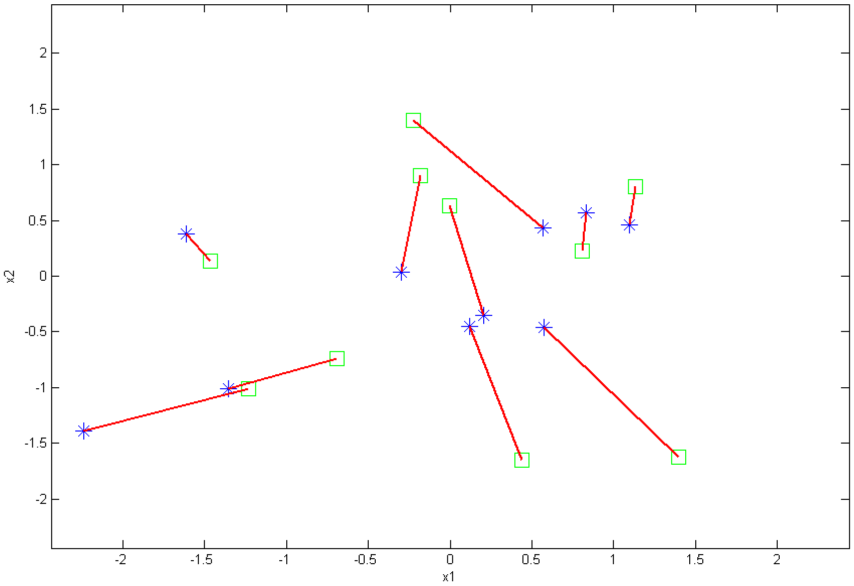

Figure 2. Minimum-weight bipartite matching on 20 points; *m*=10, *n*=10.

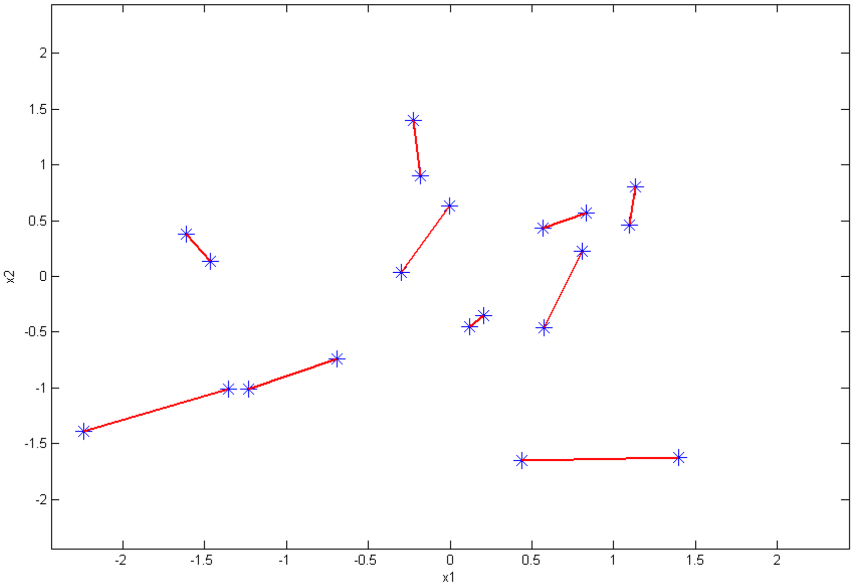

Figure 3. Minimum-weight non-bipartite matching on 20 points.

(5) The Cross-Match Statistic. With these ideas in place, we conclude this chapter by reviewing a recent non-bipartite matching-based approach to change-point detection by Rosenbaum (2005). Rosenbaum presents an exact distribution-free test to detect change in multivariate data using non-bipartite matching. We explain his test in a bit more detail than the previous tests, as his ideas have substantially motivated our thinking on this problem. In fact, one of our results extends his single non-sequential test to a simultaneous test.

Consider the case where $\mathsf{A}_1 = \{X_1, \ldots, X_m\}$ is the set of the first m observations and $\mathsf{A}_2 = \{X_{m+1}, \ldots, X_{m+n}\}$ is the set of the last $n = N - m$ observations. The goal is to test for equality of distributions for the populations associated with A_1 and A_2. Compute an optimal non-bipartite matching on $\mathsf{A}_1 \cup \mathsf{A}_2$ and let M^C denote the number of pairs cross-matched between A_1 and A_2. Rosenbaum (2005) calls M^C (which he denotes "A_1") the cross-match statistic and derives its exact null distribution. For the case where N is even, this distribution is given by

$$P\left(M^C = k\right) = \frac{2^k (N/2)!}{\binom{N}{m}\left(\frac{m-k}{2}\right)!k!\left(\frac{n-k}{2}\right)!}, \tag{2.24}$$

where k takes even values from 0 to $\min(m,n)$ if m and n are both even, k takes odd values from 1 to $\min(m,n)$ if m and n are both odd, and $P\left(M^C = k\right) = 0$ for all other k. The case where N is odd may be accommodated by introducing a pseudo-observation X_{N+1} such that $c\left(X_i, X_{N+1}\right) = 0 \ \forall i$, computing a non-bipartite matching on the pooled sample $\left\{X_1, \ldots, X_{N+1}\right\}$, and discarding the pair that includes the pseudo-observation. Equation (2.24) is then adjusted to refer to the remaining $N-1$ observations and $(N-1)/2$ pairs, conditioned on the set membership of the observation with which the pseudo-observation is paired. The distribution for odd N becomes

$$\begin{aligned}
P\left(M^C = k\right) &= \begin{bmatrix} P\left(M^C = k \middle| X_{N+1} \text{ is paired with an element of } \mathsf{A}_1\right) \\ \times P\left\{X_{N+1} \text{ is paired with an element of } \mathsf{A}_1\right\} \end{bmatrix} \\
&\quad + \begin{bmatrix} P\left(M^C = k \middle| X_{N+1} \text{ is paired with an element of } \mathsf{A}_2\right) \\ \times P\left\{X_{N+1} \text{ is paired with an element of } \mathsf{A}_2\right\} \end{bmatrix} \\
&= \frac{2^k ((N-1)/2)!}{\binom{N-1}{m-1}\left(\frac{m-k-1}{2}\right)!k!\left(\frac{n-k}{2}\right)!} \cdot \frac{m}{N} \cdot I\left(\left\{m-k \equiv 1 \bmod 2\right\}\right) \\
&\quad + \frac{2^k ((N-1)/2)!}{\binom{N-1}{m}\left(\frac{m-k}{2}\right)!k!\left(\frac{n-k-1}{2}\right)!} \cdot \frac{n}{N} \cdot I\left(\left\{m-k \equiv 0 \bmod 2\right\}\right) \\
&= \frac{2^k ((N-1)/2)!}{\binom{N}{m}k!}\left(\frac{I\left(\left\{m-k \equiv 1 \bmod 2\right\}\right)}{\left(\frac{m-k-1}{2}\right)!\left(\frac{n-k}{2}\right)!} + \frac{I\left(\left\{m-k \equiv 0 \bmod 2\right\}\right)}{\left(\frac{m-k}{2}\right)!\left(\frac{n-k-1}{2}\right)!}\right), \qquad (2.25)
\end{aligned}$$

where $I(\cdot)$ is the indicator function and the factorial terms that depend on m or n are only computed when the factorial argument is an integer (since otherwise the indicator function in the numerator is zero). In this case, k takes values from 0 to $\min(m,n)$. In any case, the null hypothesis that the distributions underlying A_1 and A_2 are equal is rejected for small values of M^C, since different underlying distributions lead to a preference for within-group matching over cross-group matching.

While (2.24) and (2.25) are exact probabilities, in practice these values can be difficult to compute for large *N*. Rosenbaum proves that under the null hypothesis, the conditional distribution of M^C given *m* converges in distribution to the normal distribution for $m / N \to p$ (constant); that is,

$$\frac{M^C - \mu_{M^C}}{\sigma_{M^C}} \xrightarrow{\mathrm{D}} N\left(0,1\right), \tag{2.26}$$

where

$$\mu_{M^C} = E\left[M^C\right] = \frac{mn}{N-1} \quad \text{and} \quad \sigma^2_{M^C} = \mathrm{Var}\left[M^C\right] = \frac{2m\left(m-1\right)n\left(n-1\right)}{\left(N-3\right)\left(N-1\right)^2} \tag{2.27}$$

for even *N*. This enables application of the cross-match test to cases where *N* is large.

We provide a small example to illustrate this test: Figure 4 shows the familiar data cloud from previous figures; in this case the data are associated with two subgroups of equal size. Group labels '□' and '*' are assigned at random to model two groups of 8 and 12 points, respectively, that are drawn from a common distribution. Cross-matches are circled; for this case Rosenbaum's cross-match statistic takes the value $M^C = 4$ and has an associated *p*-value of $P\left(M^C \leq 4\right) = 0.48$. Clearly no significant evidence exists to infer that the distributions underlying the two groups are different.

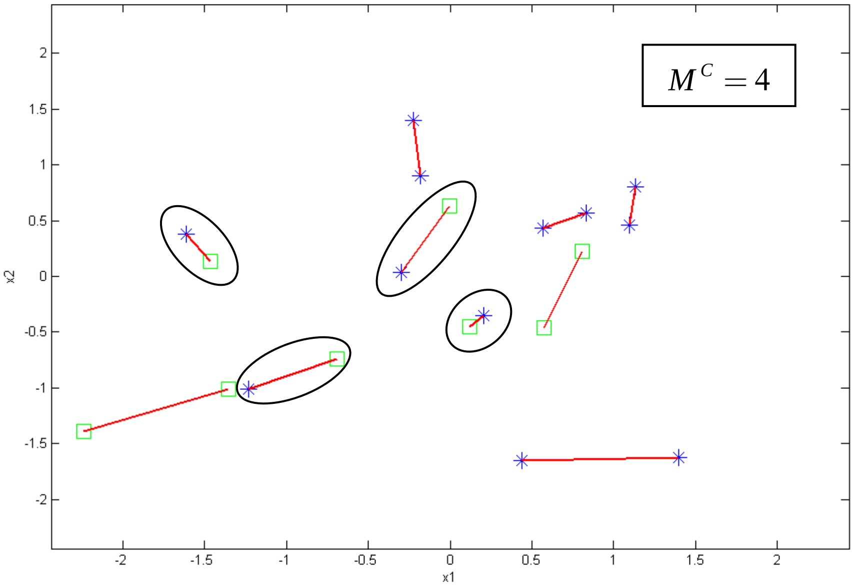

Figure 4. **Rosenbaum's cross-match statistic with no change point.**

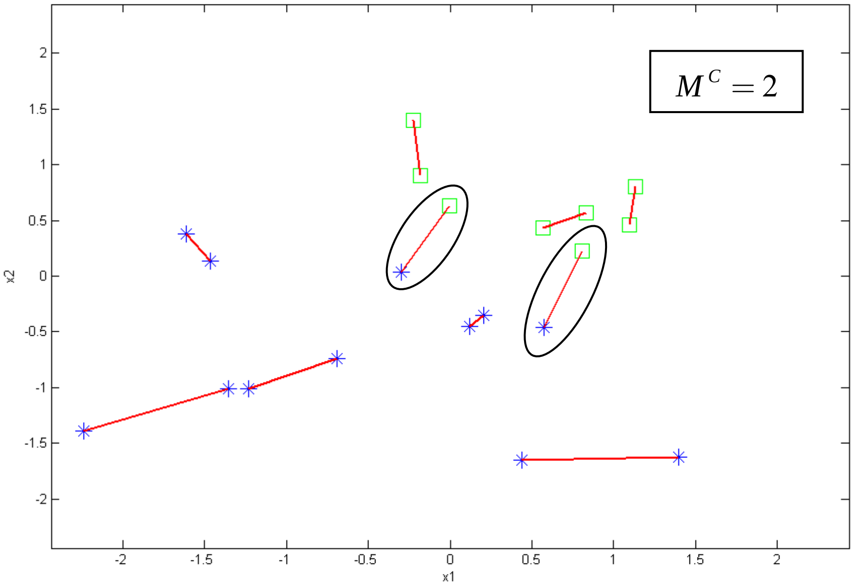

Figure 5. **Rosenbaum's cross-match statistic with change point.**

On the other hand, Figure 5 shows the same scatter plot, where this time one group consists of the lower-left hand observations ('$*$') and the other group consists of the upper-right hand observations ('$\square$'). Again, cross-matches are circled, and in this case $M^C = 2$ with an associated *p*-value of $P\left(M^C \leq 2\right) = 0.08$. This *p*-value is certainly stronger evidence that the two groups have different underlying distributions; for this small sample size the smallest achievable *p*-value is $P\left(M^C = 0\right) = 0.0014$.

Other examples of optimal non-bipartite matching techniques applied to statistical problems are found in Lu *et al.* (2001), Lu and Rosenbaum (2004), and Greevy *et al.* (2004). In an observational study of a media campaign against drug abuse, Lu *et al.* (2001) use optimal non-bipartite matching to pair teen subjects in such a way that each pair is demographically similar but has markedly different exposure to the media campaign. The evaluation compares stated intentions related to illegal drugs among comparable teens to assess the effectiveness of the campaign. Lu and Rosenbaum (2004) transform a tripartite problem of comparing one test group to two control groups into a non-bipartite matching problem in order to evaluate whether or not a localized minimum wage increase could be associated with depressed low-wage employment in that area. Greevy *et al.* (2004) demonstrate that optimal non-bipartite matching leads to improved covariate balance over randomized block design in a study of a cardiac function treatment for child cancer survivors.

The tests we propose belong to the small but growing category of graph-theoretic tests for homogeneity that involve minimizing sums of interpoint costs on graphs; this category includes Friedman and Rafsky's (1979) MST test and Rosenbaum's (2005) cross-match test. The null hypothesis of homogeneity implies that the structure of these graphs is indifferent to group labeling, which permits the derivation of null distributions by straightforward arguments. We will apply this principle to derive null distribution results for our tests.

In summary, much room remains for innovative work in the field of multivariate, nonparametric change-point detection. Particularly, very few change-point tests exist with the properties of being *multivariate*, *simultaneous*, and *distribution-free*. We proceed now to present tests with exactly these properties.

THIS PAGE INTENTIONALLY LEFT BLANK

III. THEORETICAL RESULTS

Suppose that we have an even number $N = 2n \geq 4$ observations $\{X_1, \ldots, X_N\}$ ordered with respect to time (or any other variable) and we want to test whether the observations are changing with respect to this ordering. For example, we might want to test for a jump or a drift beginning at some unknown point in the sequence. The observations may be multivariate, but are assumed to be independent. The requirement that N be even is not strictly necessary, but it does simplify the exposition (we explain later how our results extend to odd sample sizes) .

For a non-sequential simultaneous test where F_i denotes the distribution of the i^{th} sample value, the null hypothesis of homogeneity asserts that $F_1 = F_2 = \cdots = F_N$ without specifying the common distribution. The alternative hypothesis asserts that there exists an integer τ, $1 \leq \tau_0 \leq \tau \leq \tau_1 \leq N-1$, such that $F_1 = F_2 = \cdots = F_{\tau-1}$, $F_{\tau-1} \neq F_\tau$, and $\delta\left(F_1, F_j\right) - \max_{k\in\{\tau,\ldots,j\}} \delta\left(F_k, F_j\right)$ is strictly positive over $j \in \{\tau, \ldots, N\}$, usually with $\tau_0 = 1$ and $\tau_1 = N-1$ as discussed previously. With respect to a given cost function we compute an optimal non-bipartite matching $M = \left\{\left\{X_{j_1}, X_{j_2}\right\}, \ldots, \left\{X_{j_{2n-1}}, X_{j_{2n}}\right\}\right\}$. Let $R_1, R_2, \ldots, R_{2n}$ denote the sequence labels associated with each observation such that if $\left\{X_{j_{2i-1}}, X_{j_{2i}}\right\}$ is the i^{th} listed pair, then $R_{2i-1} = j_{2i-1}$ and $R_{2i} = j_{2i}$. Finally, order each individual pair as $\left(U_i, Y_i\right)$, where U_i and Y_i are the minimum and maximum ranks of the ordering variable respectively:

$$(3.1) \qquad U_i = \min\left\{R_{2i-1}, R_{2i}\right\} \text{ and } Y_i = \max\left\{R_{2i-1}, R_{2i}\right\}, \quad i \in \{1, \ldots, n\}.$$

With this setup in place, we are now ready to propose new change-point tests based on the ordered rank pairs associated with an optimal non-bipartite matching.

A. THE SUM OF PAIR-MAXIMA TEST

1. The T_N Statistic

Under the alternative hypothesis that some distribution change has occurred within a sequence of observations, one expects an optimal non-bipartite matching to pair observations that are closer together in sequence than would be the case under the null hypothesis. This suggests summing the differences $Y_i - U_i$ across all pairs and rejecting the null hypothesis if the sum is less than some critical value. We consider an equivalent test statistic T_N based on its relationship to this sum:

$$T_N = \sum_{i=1}^{n} Y_i = \frac{1}{2} n(2n+1) + \frac{1}{2} \sum_{i=1}^{n} (Y_i - U_i). \tag{3.2}$$

We call T_N the *Sum of Pair-Maxima* (SPM) test statistic, with rejection of the null hypothesis indicated by small values of this sum. We proceed to derive the mean and variance of T_N, and show that T_N has a limiting normal distribution by invoking a central limit theorem result attributed to Stein (1986).

2. Expected Value and Variance

A sequence $(Z_1, \ldots, Z_k)$ of random variables is said to be **exchangeable** if for any permutation π of indices $\{1, \ldots, k\}$, the joint probability distributions of $(Z_1, \ldots, Z_k)$ and $(Z_{\pi(1)}, \ldots, Z_{\pi(k)})$ are identical (Fristedt and Gray, 2004). Under the null hypothesis, each of the $N!$ possible assignments of sequence labels is equally likely and the random variables $(Y_1, \ldots, Y_n)$ are exchangeable. To obtain expressions for the expected value and variance of T_N we apply equations (3.3)-(3.5), which are derived in Appendix A:

$$P(Y_1 = t) = \frac{t-1}{n(2n-1)}, \quad t = 2, \ldots, 2n \ , \tag{3.3}$$

(3.4)
$$P\left(Y_2 = t \mid Y_1 = s\right) = \begin{cases} \dfrac{(t-1)(s-3)}{(s-1)(n-1)(2n-3)}, & t = 2,\ldots,s-1, \\ \dfrac{t-3}{(n-1)(2n-3)}, & t = s+1,\ldots,2n, \end{cases}$$

and

(3.5)
$$\begin{aligned} E\left[Y_1\right] &= \frac{2(2n+1)}{3}, \\ E\left[Y_1^2\right] &= \frac{(2n+1)(3n+1)}{3}, \\ E\left[Y_2\,Y_1\right] &= \frac{8(2n+1)(5n+2)}{45}, \\ \operatorname{Var}\left(Y_1\right) &= \frac{(2n+1)(n-1)}{9}, \\ \operatorname{Cov}\left(Y_1,Y_2\right) &= -\frac{4(2n+1)}{45}. \end{aligned}$$

The following are now immediate:

(3.6)
$$\begin{aligned} \mu_N &= E\left[T_N\right] = n\,E\left[Y_1\right] = \frac{2n(2n+1)}{3}, \\ \sigma_N^2 &= \operatorname{Var}\left[T_N\right] = n\operatorname{Var}\left[Y_1\right] + n(n-1)\operatorname{Cov}\left[Y_1,Y_2\right] \\ &= \frac{n(n-1)(2n+1)}{45}. \end{aligned}$$

3. Using Stein's Method to Establish Asymptotic Normality

One important result of our research is the establishment of a central limit theorem for T_N, namely that

(3.7)
$$P\left(\frac{T_N - \mu_N}{\sigma_N} \le t\right) \to \Phi(t) \text{ as } N \to \infty$$

for every $-\infty < t < \infty$, where Φ is the standard normal cumulative distribution function. While the Y_i's in the sum $T_N = \sum_{i=1}^{n} Y_i$ do have identical marginal distributions, they are not independent so the classical central limit theorem is not applicable here. Instead, we prove (3.7) by a technique referred to as *Stein's method*. Stein's method (Stein, 1972, 1986) is based on a simple differential equation that characterizes the normal distribution, and an idea called "coupling," which involves the construction of auxiliary random variables that are "close" to the variables under investigation. This method establishes bounds on the distance from normality for certain cases of dependence, including our case.

We exploit the combinatorial structure of T_N to construct an exchangeable coupling that allows us to invoke Stein's results. Let $W_N = (T_N - \mu_N)/\sigma_N$. On the same probability space on which T_N is defined we construct a random variable "close" to T_N, which we denote $\tilde{T}_N$, by selecting distinct integers u and v, $u<v$, at random from $1,2,\ldots,n$ (with $n \geq 3$), switching R_{2u} and R_{2v}, and taking the sum of the modified pair maxima. For $(T_N, \tilde{T}_N)$ to be an exchangeable pair simply means that $P(T_n = t, \tilde{T}_n = \tilde{t}) = P(T_n = \tilde{t}, \tilde{T}_n = t)$ for all t and $\tilde{t}$. This symmetry may be shown as follows. Let

(3.8)
$$T_0(u,v) = \sum_{i=1}^{u-1} Y_i + \sum_{i=u+1}^{v-1} Y_i + \sum_{i=v+1}^{n} Y_i$$

be the sum of pair-maxima for all pairs in a matching except for the u^{th} and v^{th} pair, where Y_i denotes the pair-maxima of the i^{th} pair as before. Therefore,

(3.9)
$$\begin{aligned} T_N &= T_0(u,v) + \max(R_{2u-1}, R_{2u}) + \max(R_{2v-1}, R_{2v}), \\ \tilde{T}_N &= T_0(u,v) + \max(R_{2u-1}, R_{2v}) + \max(R_{2v-1}, R_{2u}), \end{aligned}$$

for any choice of u and v. Conditioning on $T_0(u,v)$ gives

(3.10)
$$P(T_n = t, \tilde{T}_n = \tilde{t}) = \sum_{(u,v)} \sum_{t_0(u,v)} P\left(T_n = t, \tilde{T}_n = \tilde{t} \middle| T_0(u,v) = t_0(u,v)\right) P\left(T_0(u,v) = t_0(u,v)\right) \binom{n}{2}^{-1}$$

Now let $a_1(u,v) < a_2(u,v) < a_3(u,v) < a_4(u,v)$ be the ordered values of R_{2u-1}, R_{2u}, R_{2v-1}, and R_{2v}. Then, conditional on $T_0(u,v)$, the only values T_N and $\tilde{T}_N$ can assume are $b_1(u,v) = T_0(u,v) + a_2(u,v) + a_4(u,v)$ and $b_2(u,v) = T_0(u,v) + a_3(u,v) + a_4(u,v)$. It follows directly that

$$\begin{aligned}
&P\left(T_N = b_1(u,v), \tilde{T}_N = b_1(u,v) \middle| T_0(u,v)\right) = 0; \\
&P\left(T_N = b_2(u,v), \tilde{T}_N = b_1(u,v) \middle| T_0(u,v)\right) = \frac{1}{3}; \\
&P\left(T_N = b_1(u,v), \tilde{T}_N = b_2(u,v) \middle| T_0(u,v)\right) = \frac{1}{3}; \\
&P\left(T_N = b_2(u,v), \tilde{T}_N = b_2(u,v) \middle| T_0(u,v)\right) = \frac{1}{3};
\end{aligned} \tag{3.11}$$

so the joint conditional probability distribution $P\left(T_N, \tilde{T}_N \middle| T_0(u,v)\right)$ is symmetric. Therefore, by (3.10), $P\left(T_n = t, \tilde{T}_n = \tilde{t}\right) = P\left(T_n = \tilde{t}, \tilde{T}_n = t\right)$ and so $\left(T_N, \tilde{T}_N\right)$ is an exchangeable pair. Now define $\tilde{W}_N = \left(\tilde{T}_N - \mu_N\right)/\sigma_N$. In the same manner $\left(W_N, \tilde{W}_N\right)$ is an exchangeable pair.

Moreover, $\tilde{T}_N = T_N + \Delta_{N;u,v}$, where $\Delta_{N;u,v} = \tilde{Y}_u + \tilde{Y}_v - Y_u - Y_v$ with $\tilde{Y}_u = \max\left(R_{2u-1}, R_{2v}\right)$ and $\tilde{Y}_v = \max\left(R_{2u}, R_{2v-1}\right)$. Then for any $i \in \{1, \ldots, n\}$, $E\left[Y_i \mid W_N\right] = n^{-1} T_N$ and

$$E\left[\tilde{Y}_i \mid W_N\right] = \frac{Q_N}{2n(n-1)} = \frac{(2n+1)(2n-1)}{3(n-1)} - \frac{T_N}{2n(n-1)} \tag{3.12}$$

where

$$\begin{aligned}
Q_N &= \sum_{k=0}^{1} \sum_{\ell=0}^{1} \sum_{j_2=2}^{n} \sum_{j_1=1}^{j_2-1} \max\left(R_{2j_1-k}, R_{2j_2-\ell}\right) = \sum_{j=2}^{2n} \sum_{i=1}^{j-1} \max(i,j) - T_N \\
&= \sum_{j=2}^{2n} j(j-1) - T_N \\
&= \frac{2n(2n+1)(2n-1)}{3} - T_N \quad .
\end{aligned} \tag{3.13}$$

Combining these expressions gives

$$E\left[\tilde{W}_N \mid W_N\right] = \left(1-\lambda_N\right)W_N\,, \tag{3.14}$$

where

$$\lambda_N = \frac{2n-1}{n\left(n-1\right)}\,. \tag{3.15}$$

From Theorem 2.5 of Rinott and Rotar (2000) we have

$$\begin{aligned}\left|P\left(W_N \le t\right) - \Phi\left(t\right)\right| \le &\frac{6}{\lambda_N}\sqrt{\text{Var}\left\{E\left[(\tilde{W}_N - W_N)^2 \mid W_N\right]\right\}} \\ &+ \frac{6}{\sqrt{\lambda_N}}\sqrt{E\left\{\left|\tilde{W}_N - W_N\right|^3\right\}}\quad,\end{aligned} \tag{3.16}$$

which in the present case reduces to

$$\begin{aligned}\left|P\left(W_N \le t\right) - \Phi\left(t\right)\right| \le &\frac{90}{n^2}\sqrt{\text{Var}\left\{E\left[\Delta^2_{N;u,v} \mid W_N\right]\right\}} \\ &+ \frac{6\cdot 45^{3/4}}{n^{7/4}}\sqrt{E\left\{\left|\Delta_{N;u,v}\right|^3\right\}}\quad.\end{aligned} \tag{3.17}$$

We now show that $n^{-4}\text{Var}\left\{E\left[\Delta^2_{N;u,v} \mid W_N\right]\right\} \to 0$ and $n^{-7/2}E\left\{\left|\Delta_{N;u,v}\right|^3\right\} \to 0$ from which $\left|P\left(W_N \le t\right) - \Phi\left(t\right)\right| \to 0$ will follow. Because the second condition easily follows from the fact that $|\Delta_{N;u,v}| \le 2n-3$, we focus on the first condition. We cite two results that will be useful:

<u>Lemma 3-1</u>: Suppose that $\Im_1$ and $\Im_2$ are $\sigma-$fields with $\Im_1 \subseteq \Im_2$. Then for any random variable U that is measurable with respect to both $\Im_1$ and $\Im_2$

$$\text{Var}\left\{E\left[U \mid \Im_1\right]\right\} \le \text{Var}\left\{E\left[U \mid \Im_2\right]\right\}. \tag{3.18}$$

<u>Proof of Lemma 3-1</u>: See Theorem 34.4 of Billingsley (1986), p. 470. Let $V = E\left[U \mid \Im_2\right]$, giving $E\left[V \mid \Im_1\right] = E\left[U \mid \Im_1\right]$ and $\text{Var}\left\{E\left[U \mid \Im_2\right]\right\} =$ $\text{Var}(V) = \text{Var}\left\{E\left[V \mid \Im_1\right]\right\} + E\left\{\text{Var}\left[V \mid \Im_1\right]\right\} \ge \text{Var}\left\{E\left[V \mid \Im_1\right]\right\} = \text{Var}\left\{E\left[U \mid \Im_1\right]\right\}$. ■

Lemma 3-2: Let m and N be positive integers with $2m \le N$, and let (π_1, π_2) denote the first and last m elements of a random permutation of size $2m$ taken from the integers $\{1, 2, \ldots, N\}$. Then for any real-valued function g satisfying $E\left[g^2(\pi_1)\right] < \infty$,

(3.19) $$\text{Cov}\left[g(\pi_1), g(\pi_2)\right] = -\text{Cov}_1\left[g(\pi_1), g(\pi_2)\right]\left(p_{N,m}^{-1} - 1\right),$$

where $\text{Cov}_1\left[g(\pi_1), g(\pi_2)\right]$ refers to the covariance taken over randomly selected pairs of m-permutations with at least one common element, and $p_{N,m} = \frac{(N-m)!(N-m)!}{N!(N-2m)!}$ is the probability that two randomly selected pairs of m-permutations have no element in common.

Proof of Lemma 3-2: Let $E_0\left[g(\pi_1)g(\pi_2)\right] = \mu_g^2$ denote the expected value of the product $g(\pi_1)g(\pi_2)$ where the permutations π_1 and π_2 are chosen independently from $\{1, 2, \ldots, N\}$, and $E\left[g(\pi_1)\right] = \mu_g$. Then

(3.20) $$E_0\left[g(\pi_1)g(\pi_2)\right] = p_{N,m}E\left[g(\pi_1)g(\pi_2)\right] + \left(1 - p_{N,m}\right)E_1\left[g(\pi_1)g(\pi_2)\right],$$

where $E_1(\cdot)$ refers to expectation taken over pairs of permutations that have at least one element in common, so

(3.21) $$\begin{aligned} E\left[g(\pi_1)g(\pi_2)\right] &= p_{N,m}^{-1}\left(E_0\left[g(\pi_1)g(\pi_2)\right] - \left(1 - p_{N,m}\right)E_1\left[g(\pi_1)g(\pi_2)\right]\right) \\ &= p_{N,m}^{-1}\left(\mu_g^2 - \left(1 - p_{N,m}\right)E_1\left[g(\pi_1)g(\pi_2)\right]\right). \end{aligned}$$

Therefore,

(3.22) $$\begin{aligned} \text{Cov}\left[g(\pi_1), g(\pi_2)\right] &= E\left[g(\pi_1)g(\pi_2)\right] - \mu_g^2 \\ &= p_{N,m}^{-1}\left(\mu_g^2 - \left(1 - p_{N,m}\right)E_1\left[g(\pi_1)g(\pi_2)\right]\right) - \mu_g^2 \\ &= \left(E_1\left[g(\pi_1)g(\pi_2)\right] - \mu_g^2\right)\left(p_{N,m}^{-1} - 1\right) \\ &= -\text{Cov}_1\left[g(\pi_1), g(\pi_2)\right]\left(p_{N,m}^{-1} - 1\right) \end{aligned}$$

as asserted. ■

By Lemma 3-1 it is sufficient to show that $n^{-4}\mathrm{Var}\left\{E\left[\Delta^2_{N;u,v} \mid \Im_N\right]\right\} \to 0$ for $\Im_N$ = $\sigma\left(R_1,\ldots,R_N\right)$, where

$$E\left[\Delta^2_{N;u,v} \mid \Im_N\right] = 2\frac{\sum_{r=2}^{n}\sum_{s=1}^{r-1}\Delta^2_{N;r,s}}{n\left(n-1\right)}. \tag{3.23}$$

Taking the variance yields

$$\begin{aligned}\mathrm{Var}\left\{E\left[\Delta^2_{N;u,v} \mid \Im_N\right]\right\} &= 2\frac{\mathrm{Var}\left[\Delta^2_{N;1,2}\right]}{n\left(n-1\right)} + 4\left(n-2\right)\frac{\mathrm{Cov}\left(\Delta^2_{N;1,2},\Delta^2_{N;1,3}\right)}{n\left(n-1\right)} \\ &\quad + \left(n-2\right)\left(n-3\right)\frac{\mathrm{Cov}\left(\Delta^2_{N;1,2},\Delta^2_{N;3,4}\right)}{n\left(n-1\right)}.\end{aligned} \tag{3.24}$$

Now use the fact that $|\Delta_{n;u,v}| \le 2n-3$ to show that the first two terms on the right in (3.24) go to zero when multiplied by n^{-4}. By Lemma 3-2, $\left|\mathrm{Cov}\left[\Delta^2_{N;1,2},\Delta^2_{N;3,4}\right]\right| \le \left|\left(2n-3\right)^4\left(p^{-1}_{N,4}-1\right)\right|$, where

$$\begin{aligned}p^{-1}_{N,4}-1 &= \frac{\left(N-4\right)\left(N-5\right)\left(N-6\right)\left(N-7\right)}{N\left(N-1\right)\left(N-2\right)\left(N-3\right)}-1 \\ &= -\frac{8\left(2N-7\right)\left(N^2-7N+15\right)}{N\left(N-1\right)\left(N-2\right)\left(N-3\right)} \\ &= O\left(n^{-1}\right).\end{aligned} \tag{3.25}$$

It follows that $\left|P\left(W_N \le t\right)-\Phi\left(t\right)\right| \to 0$ as claimed. ■

4. An Improvement to the Normal Approximation by Edgeworth Expansion

For small or moderate *n*, the error associated with the normal approximation may be unsatisfactorily large. This error can be reduced slightly by an Edgeworth expansion, which approximates the distribution of interest using the normal distribution plus higher order corrections that adjust for non-zero moments of third order and above. We derive an Edgeworth expansion to include the skewness of T_N, $\kappa_3 = E\left[\left(T_N - \mu_N\right)^3\right]$, as follows: Using conditioning arguments as before, we have

$$E\left[Y_1^3\right] = \frac{(2n+1)\left(24n^2+15n+1\right)}{15},$$

(3.26)
$$E\left[Y_1^2 Y_2\right] = \frac{4(2n+1)\left(15n^2+10n+1\right)}{45},$$

$$E\left[Y_1 Y_2 Y_3\right] = \frac{16(2n+1)\left(70n^2+49n+6\right)}{945}.$$

See Appendix A for details. Now write

(3.27)
$$T_N^3 = \sum_{i=1}^{n} Y_i^3 + 3\sum_{i\neq j} Y_i^2 Y_j + \sum_{i\neq j\neq k} Y_i Y_j Y_k$$

and apply exchangeability to obtain

(3.28)
$$\begin{aligned} E\left[T_N^3\right] &= nE\left[Y_1^3\right] + 3n(n-1)E\left[Y_1^2 Y_2\right] + n(n-1)(n-2)E\left[Y_1 Y_2 Y_3\right] \\ &= n\frac{(2n+1)\left(24n^2+15n+1\right)}{15} + 3n(n-1)\frac{4(2n+1)\left(15n^2+10n+1\right)}{45} \\ &\quad + n(n-1)(n-2)\frac{16(2n+1)\left(70n^2+49n+6\right)}{945} \\ &= \frac{n(2n+1)}{945}\left(1120n^4+1204n^3+236n^2-43n+3\right). \end{aligned}$$

Therefore,

(3.29)
$$\begin{aligned} E\left[\left(T_N-\mu_N\right)^3\right] &= E\left[T_N^3\right] - 3\mu_N\sigma_N^2 - \mu_N^3 \\ &= \frac{-n(n-1)(2n+1)(2n+3)}{945} = \kappa_3. \end{aligned}$$

Note that $\kappa_3 < 0 \;\; \forall n > 1$, so T_N is negatively skewed for all cases (since by assumption we consider cases with at least two pairs).

Now let F_N be the distribution function for the standardized random variable $Z_N = \left(T_N - \mu_N\right)/\sigma_N$, and let $\lambda_r = E\left[Z_N^r\right]$ denote the r^{th} cumulant of Z_N. In particular, (3.30) $$\lambda_3 = E\left[Z_N^3\right] = \frac{\kappa_3}{\sigma_N^3} = \frac{-\sqrt{45}\left(2n+3\right)}{21\sqrt{n\left(n-1\right)\left(2n+1\right)}}.$$

Then the Edgeworth expansion for F_N with respect to the standard normal distribution function Φ may be written

(3.31) $$\begin{aligned} F_N\left(x\right) &= \Phi\left(x\right) - \frac{\lambda_3 \Phi'''\left(x\right)}{6\sqrt{n}} + O\left(n^{-1}\right) \\ &= \Phi\left(x\right) - \frac{\lambda_3}{6\sqrt{2\pi n}}\left(x^2-1\right)e^{-x^2/2} + O\left(n^{-1}\right) \\ &= \Phi\left(x\right) + \frac{\sqrt{45}}{126\sqrt{2\pi}} \cdot \frac{\left(2n+3\right)}{n\sqrt{\left(n-1\right)\left(2n+1\right)}}\left(x^2-1\right)e^{-x^2/2} + O\left(n^{-1}\right) \end{aligned}$$

(Wallace, 1958). It is evident that the Edgeworth expansion (3.31) makes a positive correction to the normal distribution outside one standard deviation from the mean, and a negative correction inside one standard deviation. Table 1 shows the critical values for $N = 200$ obtained by Edgeworth expansion compared to estimates obtained by simulation and normal approximation.

α	Simulation	Edgeworth	Normal
0.001	12746	12749	12750
0.005	12854	12857	12858
0.01	12908	12910	12911
0.05	13050	13054	13054
0.1	13128	13130	13131

Table 1. Critical values of the T_N statistic for *N*=200 estimated by 10,000 simulations, Edgeworth expansion, and normal approximation.

The improvement appears rather small, but taking the skewness of T_N into account in an Edgeworth expansion provides improvement nonetheless and reduces SPM test false alarm rates slightly relative to the normal approximation. Appendix B lists approximate critical values for various significance levels and sample sizes using (3.31).

5. Treatment of Odd Sample Sizes

If the sample size is odd, non-bipartite matching will leave one data point unassigned. Conceptually, it poses no difficulty to extend our results to this case as we now proceed to do. Let N denote the total sample size as before. For clarity we let T_{N0} and T_{N1} denote the sum of n matched-pair maxima in the even ($N = 2n$) and odd ($N = 2n+1$) cases respectively. Define a stochastic replica of T_{N1} as follows:

1) Select integer u at random from the set $\{1, 2, \ldots, 2n+1\}$;

2) If $u \le 2n$ take $\tilde{T}_{N1} = T_{N0} + 2n + 1 - Y_{\lfloor (u+1)/2 \rfloor}$;

3) Otherwise take $\tilde{T}_{N1} = T_{N0}$.

Equivalently, $\tilde{T}_{N1} = T_{N0} + \delta_n \left(2n + 1 - Y_{\lfloor (u+1)/2 \rfloor}\right)$ where δ_n is an independent Bernoulli random variable with success probability $2n/(2n+1)$. Using this expression, the expected value and variance are obtained:

$$\begin{aligned} \mu_{N1} &= E[T_{N1}] = E\left[\tilde{T}_{N1}\right] = E[T_{N0}] + \frac{2n}{2n+1}\left[2n+1-\frac{2(2n+1)}{3}\right] \\ &= \frac{2n(2n+1)}{3} + \frac{2n}{3} \quad = \frac{4n(n+1)}{3} \\ &= \mu_{N0}\left(\frac{2n+2}{2n+1}\right); \end{aligned} \tag{3.32}$$

$$\begin{aligned}\sigma_{N,1}^2 &= \mathrm{Var}\left[\tilde{T}_{N1}\right] = E\left[\mathrm{Var}\left[\tilde{T}_{N1} \mid \delta_n\right]\right] + \mathrm{Var}\left[E\left[\tilde{T}_{N1} \mid \delta_n\right]\right] \\ &= \frac{n(n-1)(2n+7)}{45} + \frac{2n}{9} = \frac{n(n+1)(2n+3)}{45} \\ &= \sigma_{N0}^2 \cdot \frac{(n+1)(2n+3)}{(n-1)(2n+1)} \quad .\end{aligned} \tag{3.33}$$

In standardized terms, the correction made by adding an observation to the even case becomes negligible as the sample size increases and does not affect the asymptotic normality argument of the previous section.

6. On the Consistency of T_N

Henze and Penrose (1999) establish that Friedman and Rafsky's minimum spanning tree test is consistent against all alternatives for the two sample case $X_1,\ldots,X_m \sim F$, $X_{m+1},\ldots,X_{m+n} \sim G$, $H_0: F = G$, $H_1: F \neq G$, F and G unknown. Rosenbaum (2005) argues for the consistency of his cross-match statistic M^C distributed as in (2.24) and (2.25) against alternatives of the form $X_1,\ldots,X_m \sim F \neq G \sim X_{m+1},\ldots,X_{m+n}$ by showing that it is a consistent test for comparing two *discrete* distributions with *finitely* many mass points. He heuristically extends that consistency argument to the general case by the fact that any two distributions may be approximated arbitrarily closely by two discrete distributions with finitely many mass points. We do not try to formalize that argument here; rather we theoretically motivate the use of the SPM test under alternative hypotheses by proving a consistency result for T_N that depends on the consistency of M^C:

Proposition 3-1: T_N is consistent against all alternatives of the form

$$X_1,\ldots,X_{m-1} \sim F \neq G \sim X_m,\ldots,X_N \;\; \exists m \in \{2,\ldots,N\} \tag{3.34}$$

against which M^C *is consistent,* where M^C is Rosenbaum's cross-match statistic as defined in (2.24).

We prove the case for even N and a change point that divides the observations into two subsets each containing an even number of elements; the same reasoning extends to odd cases. Let $N = 2n$, suppose $M_1 + 1$ is a change point, M_1 and $M_2 = N - M_1$ are both even, and $M_1 / N \to \pi_1$ constant as $N \to \infty$. Adopt the subscript notation "0" or "1" to denote probabilities (or expectations or variances) taken under the null or alternative hypothesis, respectively. We will show that $P_1\left(\frac{T_N - \mu_N}{\sigma_N} < z_\alpha\right) \to 1$ as $N \to \infty$ for any significance level α, where z_α denotes that α-quantile of the standard normal distribution. We build our proof on the following fact:

Lemma 3-3: Let $T_{N|k}$ denote the random variable obtained by matching among the first M_1 points alone (call that set of pairs P_1), matching among the last M_2 points alone (call that set of pairs P_2), randomly choosing k pairs in P_1 and k pairs in P_2 (k is even for this case since M_1, M_2, and N are all even), and randomly swapping one element from each selected pair in P_1 with one element from each selected pair in P_2 (each pair gets one swap), and finally computing the pair-maxima sum on the new pairs. Let Δ_i be the change in the pair-maxima sum due to the i^{th} swap, $i = 1, \dots, k$. Then under null or alternative hypotheses,

(3.35) $T_{N|k}$ is distributed the same as T_N conditional on $M^C = 2k$

and

$$T_{N|k} \sim T_{M_1} + T_{M_2} + \frac{M_1 M_2}{2} + \sum_{i=1}^{k} \Delta_i \,, \tag{3.36}$$

where $k \in \{0, \dots, \min(M_1, M_2)/2\}$, the Δ_i are exchangeable, and each Δ_i is independent of M^C.

Proof of Lemma 3-3: That $T_{N|k}$ is distributed the same as T_N conditional on $M^C = 2k$ is a consequence of all within group matchings being equally likely and all cross-group matchings being equally likely. Each swap constitutes two cross-matches. $T_{N|k} \sim T_{M_1} + T_{M_2} + \frac{M_1 M_2}{2} + \sum_{i=1}^{k} \Delta_i$ results from the fact that by the construction of $T_{N|k}$, before any swaps occur the total pair-maxima sum is $T_{M_1}+T_{M_2}+M_1\frac{M_2}{2}$, where T_{M_1} is the pair-maxima sum over P_1 and $T_{M_2}+M_1\frac{M_2}{2}$ is the pair-maxima sum over P_2. After the swaps the final pair-maxima sum increases by a total $\sum_{i=1}^{k}\Delta_i$ relative to the pre-swap sum. Exchangeability and independence are by construction of $T_{N|k}$. ■

Proof of Proposition 3-1: By (3.36),

$$
\begin{aligned}
E_0\left[T_N\right] &= E_0\left[E_0\left[T_N \middle| M^C\right]\right] = E_0\left[E_0\left[T_{M_1}+T_{M_2}+M_1\frac{M_2}{2}+\sum_{i=1}^{k}\Delta_i \middle| M^C=2k\right]\right] \\
&= E_0\left[T_{M_1}\right]+E_0\left[T_{M_2}\right]+\frac{M_1M_2}{2}+\frac{E_0\left[M^C\right]}{2}E_0\left[\Delta_1\right].
\end{aligned}
\tag{3.37}
$$

Now solve for $E_0\left[\Delta_1\right]$ directly by substitution using (2.27) and (3.6):

(3.38)

$$
\begin{aligned}
E_0\left[\Delta_1\right] &= \frac{2\left(E_0\left[T_N\right]-E_0\left[T_{M_1}\right]-E_0\left[T_{M_2}\right]\right)-M_1M_2}{E_0\left[M^C\right]} \\
&= \frac{2\left(N\left(N+1\right)-M_1\left(M_1+1\right)-M_2\left(M_2+1\right)\right)-3M_1M_2}{3M_1M_2}\left(N-1\right) \\
&= \frac{2\left(N\left(N+1\right)-\pi_1N\left(\pi_1N+1\right)-\left(1-\pi_1\right)N\left(\left(1-\pi_1\right)N+1\right)\right)-3\pi_1\left(1-\pi_1\right)N^2}{3\pi_1\left(1-\pi_1\right)N^2}\left(N-1\right) \\
&= \frac{N-1}{3}.
\end{aligned}
$$

Alternately, $E_0\left[\Delta_1\right]$ can be found more directly by noting that with every swap the pair-maxima value from P_1 gets replaced by the pair-minima value from P_2. Denoting the pair minima of the first swapped pair in P_2 as $U_{1;2}$ and the pair maxima of the first swapped pair in P_1 as $Y_{1;1}$, we have

$$
E_0\left[\Delta_1\right]=E_0\left[U_{1;2}-Y_{1;1}\right]=\left(M_1+\frac{\left(M_2+1\right)}{3}\right)-\frac{2\left(M_1+1\right)}{3}=\frac{N-1}{3}.
\tag{3.39}
$$

Since equation (3.36) holds under both null and alternative hypotheses, we proceed as above to find an expression for $E_1\left[T_N\right]$:

$$E_1\left[T_N\right]=E_1\left[E_1\left[T_N\ \middle|M^C\right]\right]=E_1\left[E_0\left[T_{M_1}+T_{M_2}+M_1\frac{M_2}{2}+\sum_{i=1}^{k}\Delta_i\middle|M^{C\prime}=2k\right]\right]$$

$$(3.40)\qquad =E_0\left[T_{M_1}\right]+E_0\left[T_{M_2}\right]+\frac{M_1M_2}{2}+\frac{E_1\left[M^C\right]}{2}E_0\left[\Delta_1\right]$$

$$=E_0\left[T_{M_1}\right]+E_0\left[T_{M_2}\right]+\frac{M_1M_2}{2}+\left(\frac{N-1}{6}\right)E_1\left[M^C\right],$$

so

$$(3.41)\ E_1\left[T_N\right]-E_0\left[T_N\right]=\left(\frac{N-1}{6}\right)\left(E_1\left[M^{C\prime}\right]-E_0\left[M^C\right]\right)=N\left(\frac{N-1}{6}\right)\frac{E_1\left[M^C\right]-E_0\left[M^C\right]}{N}.$$

Rosenbaum argues that $\frac{E_1\left[M^C\right]-E_0\left[M^C\right]}{N}\to\delta<0$ as $N\to\infty$; therefore, for sufficiently large *N*

$$(3.42)\ \frac{E_1\left[T_N\right]-E_0\left[T_N\right]}{\sigma_N}<\frac{N\left(N-1\right)\left(\delta/2\right)}{6\sigma_N}=\frac{\sqrt{5}\,N\left(N-1\right)\delta}{\sqrt{N\left(N-2\right)\left(N+1\right)}}\to-\infty\quad\text{as }N\to\infty\ ,$$

again using (3.6). Now we need one final lemma to complete our proof:

Lemma 3-4:

$$(3.43)\qquad \frac{\mathrm{Var}_1\left[T_N\right]}{\mathrm{Var}_0\left[T_N\right]}=O\left(1\right),$$

where $O\left(\cdot\right)$ is the standard Landau notation.

Proof of Lemma 3-4: We rely on two facts that follow directly from Rosenbaum's argument for the consistency for his cross-match statistic; namely, $E_1\left[M^C\right]=O\left(N\right)$ and $\mathrm{Var}_1\left[M^C\right]=O\left(N\right)$. First, expand $\mathrm{Var}_1\left[T_N\right]$ by conditioning:

$$(3.44)\qquad \mathrm{Var}_1\left[T_N\right]=\mathrm{Var}_1\left[E_1\left[T_N\ \middle|M^C\right]\right]+E_1\left[\mathrm{Var}_1\left[T_N\ \middle|M^C\ \right]\right].$$

Now express each term on the right hand side using (3.36). The first term is simply

$$\begin{aligned}\operatorname{Var}_1\left[E_1\left[T_N \;\middle|M^C\right]\right] &= \operatorname{Var}_1\left[E_0\left[T_{M_1}+T_{M_2}+M_1\frac{M_2}{2}+\sum_{i=1}^{k}\Delta_i\middle|M^C=2k\right]\right]\\ &= \operatorname{Var}_1\left[E_0\left[T_{M_1}\right]+E_0\left[T_{M_2}\right]+\frac{M_1M_2}{2}+\frac{M^C}{2}O\left(N\right)\right]\\ &= O\left(N^2\right)\operatorname{Var}_1\left[M^C\right]\\ &= O\left(N^3\right).\end{aligned}\tag{3.45}$$

For the second term, we first note that T_{M_1} and T_{M_2} are independent by construction, so

$$\begin{aligned}E_1\left[\operatorname{Var}_1\left[T_N \;\middle|M^C\right]\right] &= E_1\left[\operatorname{Var}_0\left[T_{M_1}+T_{M_2}+M_1\frac{M_2}{2}+\sum_{i=1}^{k}\Delta_i\middle|M^{C\prime}=2k\right]\right]\\ &= E_1\left\{\operatorname{Var}_0\left[T_{M_1}\right]+\operatorname{Var}_0\left[T_{M_2}\right]+\operatorname{Var}_0\left[\sum_{i=1}^{k}\Delta_i\middle|M^{C\prime}=2k\right]\right.\\ &\qquad\left.+\operatorname{Cov}_0\left[T_{M_1}+T_{M_2},\sum_{i=1}^{k}\Delta_i\middle|M^{C\prime}=2k\right]\right\}.\end{aligned}\tag{3.46}$$

It is clear that the exchangeable Δ_i s are negatively correlated, since $\Delta_i = U_{i;2}-Y_{i;1}$ with the $U_{i;2}$ negatively correlated, the $Y_{i;1}$ negatively correlated, and the $U_{i;2}$ independent of the $Y_{i;1}$'s. Therefore,

$$\begin{aligned}\operatorname{Var}_0\left[\sum_{i=1}^{k}\Delta_i\middle|M^C=2k\right] &= \sum_{i=1}^{k}\operatorname{Var}_0\left[\Delta_i\middle|M^C=2k\right]+\sum_{i\neq j}^{k}\operatorname{Cov}_0\left[\Delta_i,\Delta_j\middle|M^{C\prime}=2k\right]\\ &= \left(\frac{M^{C\prime}}{2}\right)\operatorname{Var}_0\left[\Delta_1\right]+\frac{M^{C\prime}\left(M^{C\prime}-1\right)}{4}\operatorname{Cov}_0\left[\Delta_1,\Delta_2\right]\\ &\leq \left(\frac{M^C}{2}\right)\operatorname{Var}_0\left[U_{1;2}-Y_{1;1}\right]=M^C O\left(N^2\right).\end{aligned}\tag{3.47}$$

Furthermore, the Δ_i s can be seen to be negatively correlated with T_{M_1} and T_{M_2}, since larger T_{M_1} values are associated with larger $Y_{i;1}$ values and larger T_{M_2} values are associated with smaller $U_{i;2}$ values. So,

$$
\begin{aligned}
E_1\left[\mathrm{Var}_1\left[T_N \;\middle|M^C\right]\right] &\le E_1\left[\mathrm{Var}_0\left[T_{M_1}\right] + \mathrm{Var}_0\left[T_{M_2}\right] + \mathrm{Var}_0\left[\sum_{i=1}^{k}\Delta_i \middle| M^{C\prime} = 2k\right]\right] \\
&= E_1\left[O\left(N^3\right) + M^C O\left(N^2\right)\right] \\
&= O\left(N^3\right).
\end{aligned}
\tag{3.48}
$$

Combining (3.45) and (3.48) results in $\mathrm{Var}\left[T_N'\right] = O\left(N^3\right)$, and so $\mathrm{Var}_1\left[T_N\right] / \mathrm{Var}_0\left[T_N\right] = O(1)$ as claimed.

Finally, to conclude the proof of our consistency proposition, we apply Chebyshev's inequality. Choose any real number $s > 0$ and any significance level α. Then

$$
\begin{aligned}
\frac{1}{s^2} &\ge P_1\left(\left|T_N - E_1\left[T_N\right]\right| \ge s\sqrt{\mathrm{Var}_1\left[T_N\right]}\right) \ge P\left(T_N - E_1\left[T_N\right] \ge s\sqrt{\mathrm{Var}_1\left[T_N\right]}\right) \\
&= P_1\left(\frac{T_N - E_0\left[T_N\right]}{\sqrt{\mathrm{Var}_0\left[T_N\right]}} \ge \frac{s\sqrt{\mathrm{Var}_1\left[T_N\right]} + E_1\left[T_N\right] - E_0\left[T_N\right]}{\sqrt{\mathrm{Var}_0\left[T_N\right]}}\right) \\
&\ge P_1\left(\frac{T_N - E_0\left[T_N\right]}{\sqrt{\mathrm{Var}_0\left[T_N\right]}} \ge z_\alpha\right) \quad \text{for sufficiently large } N,
\end{aligned}
\tag{3.49}
$$

since $\mathrm{Var}_1\left[T_N\right] / \mathrm{Var}_1\left[T_N\right] = O(1)$ by Lemma 3-4 and $\left(E_1\left[T_N\right] - E_0\left[T_N\right]\right) / \sigma_N \to -\infty$ as $N \to \infty$ by (3.42). The inequality (3.49) is true for any $s > 0$, so $P_1\left(\dfrac{T_N - \mu_N}{\sigma_N} < z_\alpha\right) \to 1$ as $N \to \infty$ and the proposition holds. ■

7. A Graphical Example

We give a graphical example to illustrate how the Sum of Pair-Maxima test works. Consider the same set of 20 points drawn from a bivariate normal distribution used for illustration in Chapter II (see Appendix D for data values). Now suppose that associated with each observation is some sequence label, for example, the time order of the observation. Figures 6 and 7 show two such cases, where the plot symbol for each data point is its sequence label, and the data are paired with respect to the optimal non-bipartite matching on the set (compare to Figure 3 in the previous chapter, which is the same plot except without sequence labels). To represent a sample whose underlying distribution has not changed with respect to sequence label, the sequence labels are assigned at random in Figure 6. To represent a sample whose underlying distribution has changed with respect to sequence label, the sequence labels are assigned from lower-left to upper-right in Figure 7.

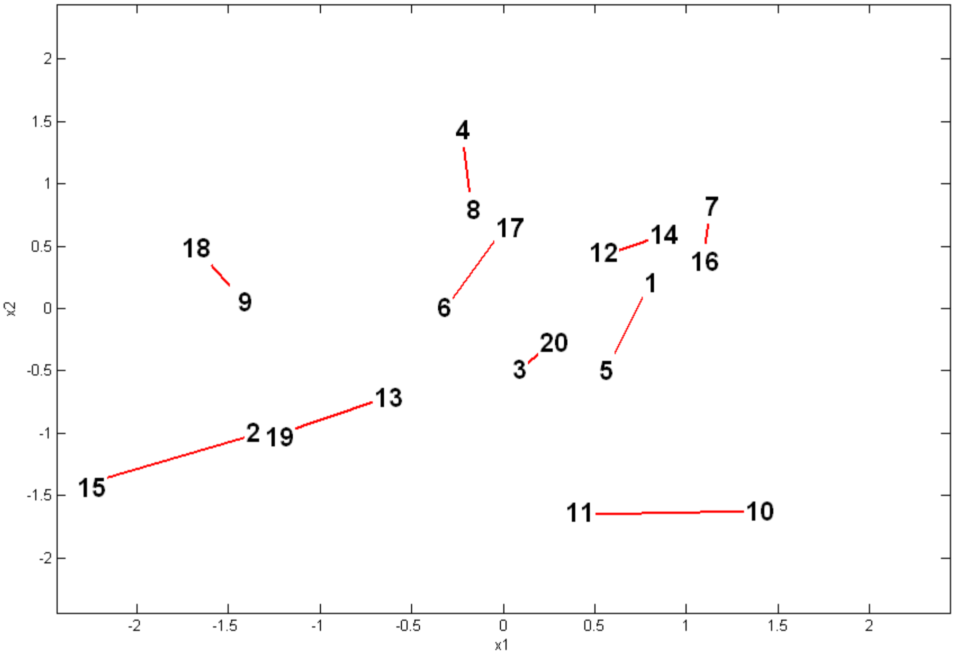

Figure 6. Minimum weight non-bipartite matching on 20 points with no change in underlying distribution with respect to observation order.

Now we compute the SPM test statistic, T_{20}, for each case and consider whether or not this test rejects the null hypothesis at significance level $\alpha = 0.05$. For $N = 20$, the quantile table in Appendix B gives $T_{20}^{\text{crit}} = 129$. In the case of Figure 6, one readily computes $T_{20} = 15+19+18+11+20+17+8+5+14+16 = 143 \geq 129 = T_{20}^{\text{crit}}$, and so the null hypothesis is not rejected. In the case of Figure 7,

$T_{20} = 2+4+6+10+11+13+17+16+18+20 = 117 < 129 = T_{20}^{\text{crit}}$, so the null hypothesis is rejected and we conclude that the underlying distribution is changing with respect to observation order. The nature of change in this example is perhaps extreme, but it serves to illustrate the sense of the T_N statistic as a change-point test.

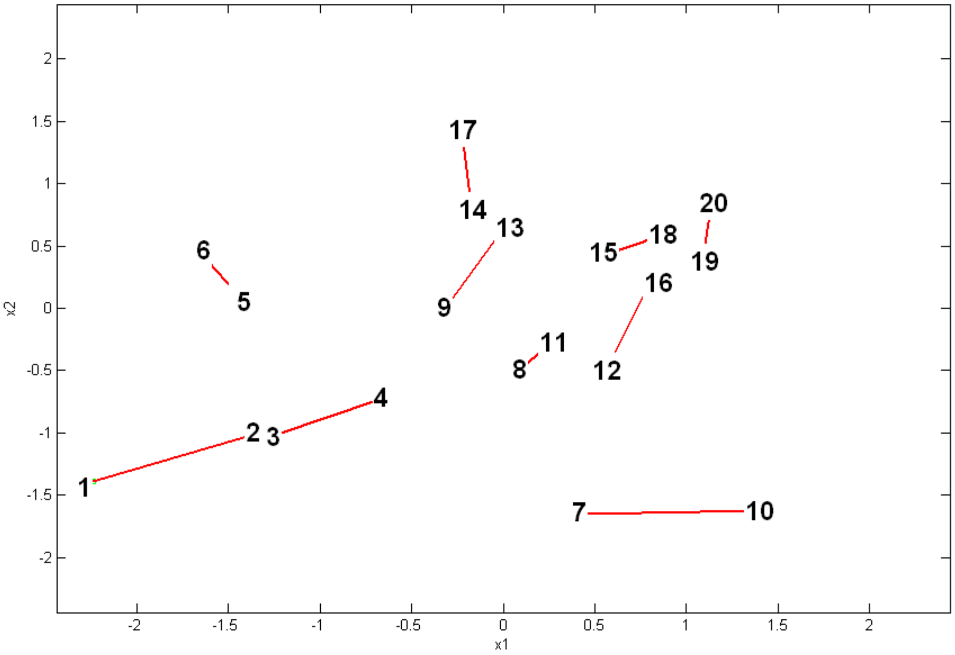

Figure 7. Minimum weight non-bipartite matching on 20 points with a change in underlying distribution with respect to observation order.

We more thoroughly examine the performance of the SPM test by simulation study in Chapter IV. As one might expect, while this nonparametric test has a fixed false alarm rate regardless of underlying distribution, its power is somewhat low. We find that we can dramatically increase the power of this test by considering a particular *ensemble* of such matching statistics, and still retain a fixed false alarm rate. Before we present the theory for such an ensemble statistic (in Section C of this chapter), we first introduce an alternative to the SPM test that is also based on non-bipartite matching.

B. THE NON-BIPARTITE ACCUMULATED PAIRS TEST

1. The $\mathbf{M}_N$ Statistic

The cross-match test statistic proposed by Rosenbaum (2005) is used when there is a clear delineation between two groups in a sample (for example, treatment and control) and the objective is to test whether the two groups come from the same probability distribution. After performing an optimal non-bipartite matching, one counts the number of pairs that link across the two predetermined groups. Under the null hypothesis of homogeneity this test statistic has a simple, exact distribution (see (2.24) and (2.25)) that can be approximated by a normal distribution in large samples. We now consider an extension of Rosenbaum's test to the situation where a sample is ordered sequentially but no prior subdivision of the observation set exists.

As before, suppose we have $N = 2n$ observations and an associated optimal non-bipartite matching $M = \{(U_1, Y_1), \ldots, (U_n, Y_n)\}$. Let $M_{k,N} = \left|\{(U_i, Y_i) \mid Y_i \le k; i = 1, \ldots, n\}\right|$ denote the number of pairings in a non-bipartite matching that occur within the first k observations, $2 \le k \le N-1$. The cross-match statistic M^C is equivalent to $M_{k,N}$, and so an exact expression for the probability mass function of $M_{k,N}$ under the null hypothesis follows directly from (2.24):

$$(3.50) \qquad P\left(M_{k,N} = r\right) = 2^{k-2r} \frac{\binom{n}{k-r}\binom{k-r}{r}}{\binom{N}{k}}, \quad r = 0 \vee (k-n), \ldots, \lfloor k/2 \rfloor,$$

where $P\left(M_{k,N} = r\right) = P\left(M^C = k - 2r\right)$ for observation subsets of size k and $N-k$. Observe that $M_{k,N}$ can be expressed in terms of the matched-pair maxima Y_j by noting that the event $M_{k,N} > r$ is identical to the event $\left|\{ j \mid Y_j \le k\}\right| > r$, which in turn is identical to the event $Y_{(r+1)} \le k$ where $Y_{(j)}$ is the j^{th} largest among the matched-pair maxima (taking $Y_{(r+1)} \equiv N+1$ for $r \ge n$). Thus, large values $M_{k,N}$ are associated with small values of $Y_{n,j}$ and vice-versa.

Rosenbaum uses the number of cross-matches as a test statistic, rejecting the null hypothesis of homogeneity for small values. The number of within-group matches in any one of the two groups, $M_{k,N}$, is an equivalent test statistic, with large values indicating evidence against the null hypothesis. We call the vector $\mathbf{M}_N = \left(M_{2,N}, \dots, M_{N-1,N}\right)'$ the *Non-Bipartite Accumulated Pairs* (NAP) test statistic. Like T_N, a test based on $\mathbf{M}_N$ rejects the null hypothesis for small values of Y_j.

2. Critical Envelope

It is possible to develop an exact simultaneous test based on $\mathbf{M}_N$ for cases where the change point k is not pre-specified. To do this we seek a vector of non-negative integers $\mathbf{q}_{N,\alpha} = \left(q_{k_0,N}, \dots, q_{k_1,N}\right)'$ (where we omit the test level subscript on the right-hand side for ease of notation below) so that the following is true for a given test level α:

(3.51) $$P\left(M_{k,N} \le q_{k,N},\ k_0 \le k \le k_1\right) \ge 1-\alpha .$$

We choose $q_{k,N}$ to be the $1-\tilde{\alpha}$ quantile of the distribution of $M_{k,N}$ so that the non-simultaneous test at stage k has level $\tilde{\alpha}$; that is, $P\left(M_{k,N} > q_{k,N} \text{ for some } k_0 \le k \le k_1\right) \le \tilde{\alpha}$. The problem, then, is how to select $\tilde{\alpha}$ so that the simultaneous test level comes as close to α as possible without exceeding this value.

To find $P\left(M_{k,N} \le q_{k,N},\ k_0 \le k \le k_1\right)$, we develop a recursive computational scheme based on the fact that

(3.52) $$P\left(M_{k,N} \le q_{k,N},\ k_0 \le k \le k_1\right) = \sum_{r=0}^{q_{k_1,N}} \pi\left(r; k_1, N\right) \cdot P\left(M_{k_1,N} = r\right),$$

where $\pi(r;k,N) = P(M_{j,N} \le q_{j,N},\ k_0 \le j \le k-1 \mid M_{k,N} = r)$, and that $\pi(r;k,N)$ has a recursive form stated in the following lemma.

<u>Lemma 3-5</u>:

(3.53)
$$\begin{aligned}\pi(r;k,N) &= \frac{2r}{k}\pi(r-1;k-1,N)\, I(r-1 \le q_{k-1,N})\\ &+\frac{k-2r}{k}\pi(r;k-1,N)\, I(r \le q_{k-1,N}),\\ &k = k_0+1,\ldots,k_1;\, r = 0 \vee (k-n),\ldots,q_{k,N}\ ,\end{aligned}$$

where $I(\cdot)$ is the indicator function.

<u>Proof of Lemma 3-5</u>: Expand $\pi(r;k,N)$ as follows:

(3.54)
$$\begin{aligned}&\pi(r;k,N)\\ &= \sum_{s=0\vee(k-1-N)}^{\lfloor (k-1)/2 \rfloor} P(M_{j,N} \le q_{j,N}, k_0 \le j \le k-2 \mid M_{k-1,N} = s, M_{k,N} = r)\cdot P(M_{k-1,N} = s \mid M_{k,N} = r)\\ &= \sum_{s=0\vee(k-1-N)}^{\lfloor (k-1)/2 \rfloor} P(M_{j,N} \le q_{j,N}, k_0 \le j \le k-2 \mid M_{k-1,N} = s)\cdot P(M_{k-1,N} = s \mid M_{k,N} = r)\\ &= \sum_{s=0\vee(k-1-N)}^{\lfloor (k-1)/2 \rfloor} \pi(s;k-1,N)\cdot P(M_{k-1,N} = s \mid M_{k,N} = r)\cdot I(s \le q_{k-1,N})\\ &= \pi(r-1;k-1,N)\cdot P(M_{k-1,N} = r-1 \mid M_{k,N} = r)\cdot I(r-1 \le q_{k-1,N})\\ &\quad + \pi(r;k-1,N)\cdot P(M_{k-1,N} = r \mid M_{k,N} = r)\cdot I(r \le q_{k-1,N})\end{aligned}$$

Finally, we note that

(3.55)

$$
\begin{aligned}
&P\left(M_{k-1,N}=r-1 \mid M_{k,N}=r\right) \\
&\quad=\frac{\sum_{s=1}^{N} P\left(M_{k-1,N}=r-1, M_{k,N}=r \mid \text{item } k \text{ matched to } s\right) \cdot P\left(\text{item } k \text{ matched to } s\right)}{P\left(M_{k,N}=r\right)} \\
&\quad=\frac{\sum_{s=1}^{k-1} P\left(M_{k-1,N}=r-1 \mid \text{item } k \text{ matched to } s\right) \cdot P\left(\text{item } k \text{ matched to } s\right)}{P\left(M_{k,N}=r\right)} \\
&\quad=\frac{k-1}{N-1} \cdot \frac{P\left(M_{k-2,N-2}=r-1\right)}{P\left(M_{k,N}=r\right)}=\frac{2r}{k}
\end{aligned}
$$

to complete the proof. ■

To start the recursion take $\pi\left(r;k_0,N\right)\equiv 1,\ \forall r$. Let

$$b\left(r;k,N\right)=\pi\left(r;k,N\right)I\left(r\le q_{k,N}\right) \tag{3.56}$$

and

$$\Delta_b\left(r;k,N\right)=b\left(r;k,N\right)-b\left(r-1;k,N\right) \tag{3.57}$$

From (3.53) we then have

$$\pi\left(r;k,N\right)=b\left(r;k-1,N\right)-\frac{2r}{k}\Delta_b\left(r;k-1,N\right) \tag{3.58}$$

which is suitable for efficient implementation in S-PLUS® (2005), R (2005), MATLAB® (2008), or other interpreted languages. There is no need for asymptotic approximations; finding an exact critical region can be done quickly at any practical sample size using trial and error. An implementation for R is included in Appendix C. We use the fact here that all information carried through conditioning to the joint events $M_{k-1,N}=s, M_{k,N}=r$ is carried through the single event $M_{k-1,N}=s$, and that the event $M_{k,N}=r$ implies either $M_{k-1,N}=r$ or $M_{k-1,N}=r-1$.

Using the exact method presented here is far less conservative than the Bonferroni method. At sample size $N = 20$ a nominal $\alpha = .05$ simultaneous test achieves test level .046 using $\tilde{\alpha} = .021$ while the Bonferroni test achieves level .0044 using $\tilde{\alpha} = .0028$ (.05 divided by 18 based on $k_0 = 2$ and $k_1 = 19$). Sample size $N = 100$ achieves level .048 using $\tilde{\alpha} = .0046$ while the Bonferroni test achieves level .006 using $\tilde{\alpha} = .0005$ (.05 divided by 98 based on $k_0 = 2$ and $k_1 = 99$).

As an illustration, Figure 8 shows two cases for $N = 100$ at significance level $\alpha = 0.05$. The solid line is the critical envelope $\mathbf{q}_{N,\alpha}$ obtained by recursion using (3.51)-(3.58). The null hypothesis case (homogeneity) is modeled by drawing all 100 points from a bivariate normal distribution $\text{BVN}(\boldsymbol{\mu}_0, \Sigma)$ where $\boldsymbol{\mu}_0 = (0,0)'$ and Σ is the identity matrix. The alternate hypothesis case (heterogeneity) is modeled by drawing the first 50 points from $\text{BVN}(\boldsymbol{\mu}_0, \Sigma)$ and the last 50 points from $\text{BVN}(\boldsymbol{\mu}_1, \Sigma)$ where $\boldsymbol{\mu}_1 = (3,0)'$. We choose a large mean jump for this example to emphasize the response of $\mathbf{M}_N$ to a change point. Applying the NAP test, we reject the null hypothesis for any case where $M_{k,N} > q_{k,N}$. In this example, $\mathbf{M}_N$ for the case of homogeneity never exceeds $\mathbf{q}_{N,\alpha}$ and so we do not reject the null hypothesis. In contrast, $\mathbf{M}_N$ for the case of heterogeneity exceeds $\mathbf{q}_{N,\alpha}$ not just once but for numerous values of *k*, so ample evidence exists to reject the null hypothesis.

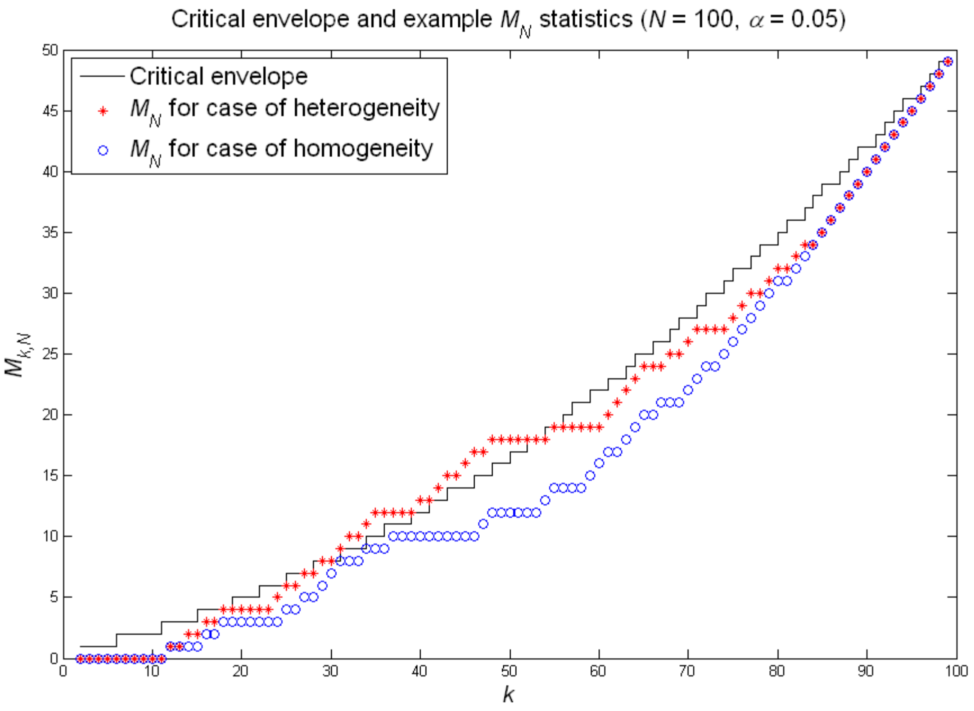

Figure 8. **Critical envelope for the NAP test and two cases of $\mathbf{M}_N$ with $N = 200$ and $\alpha = 0.05$. Reject the null hypothesis if $\mathbf{M}_N$ exceeds $\mathbf{q}_N^\alpha$ for some k.**

3. A Graphical Example

As we did for the SPM test, we illustrate the mechanics of the NAP test with the data presented in Figures 6 and 7. We compute the NAP test statistic $\mathbf{M}_{20}$ for $k = 2,\ldots,19$ and consider whether or not this test rejects the null hypothesis at significance level $\alpha = 0.05$. Figure 9 shows critical envelope $\mathbf{q}_{20,\alpha}$ and $\mathbf{M}_{20}$ for the data from Figure 6. We find the $M_{k,20}$ values easily by visual inspection: there are no pairings that occur within the first 2 observations, none in the first 3 observations, none in the first 4 observations, 1 pairing within the first 5 observations, and so on.

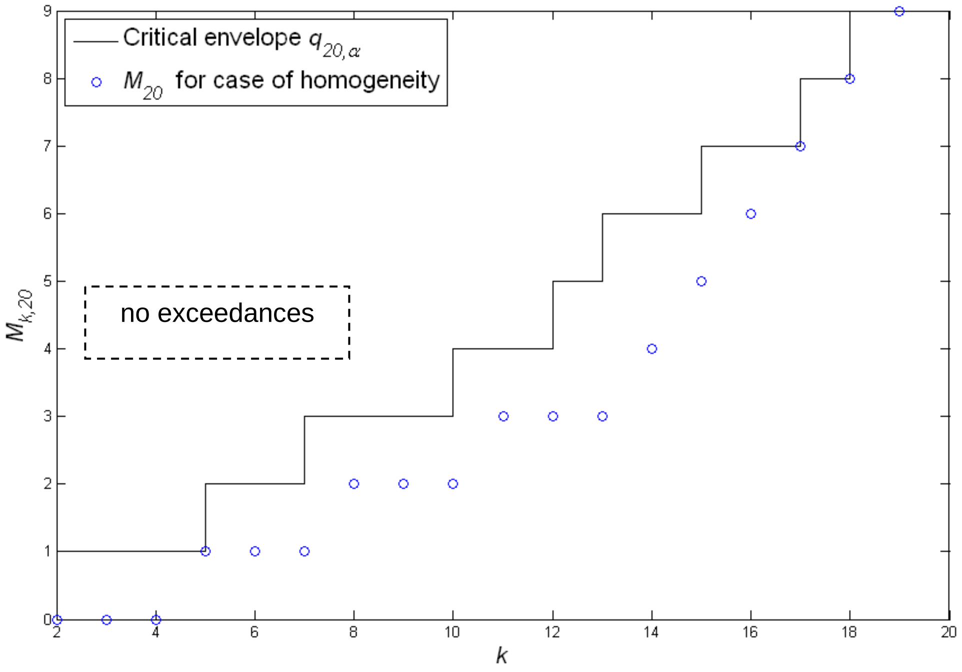

Figure 9. NAP test statistic for Figure 6 data, no change point detected.

Since $M_{k,20} \leq q_{k,20} \; \forall k$ in this case, we do not reject the null hypothesis. In contrast, Figure 10 shows $\mathbf{q}_{20,\alpha}$ and test statistic $\mathbf{M}_{20}$ for the data from Figure 7. Here we see that $M_{k,20} > q_{k,20}$ for $k = 4, 6,$ and 11. A single exceedance alone is sufficient to reject the null hypothesis, so in this case there is more than enough evidence to do so.

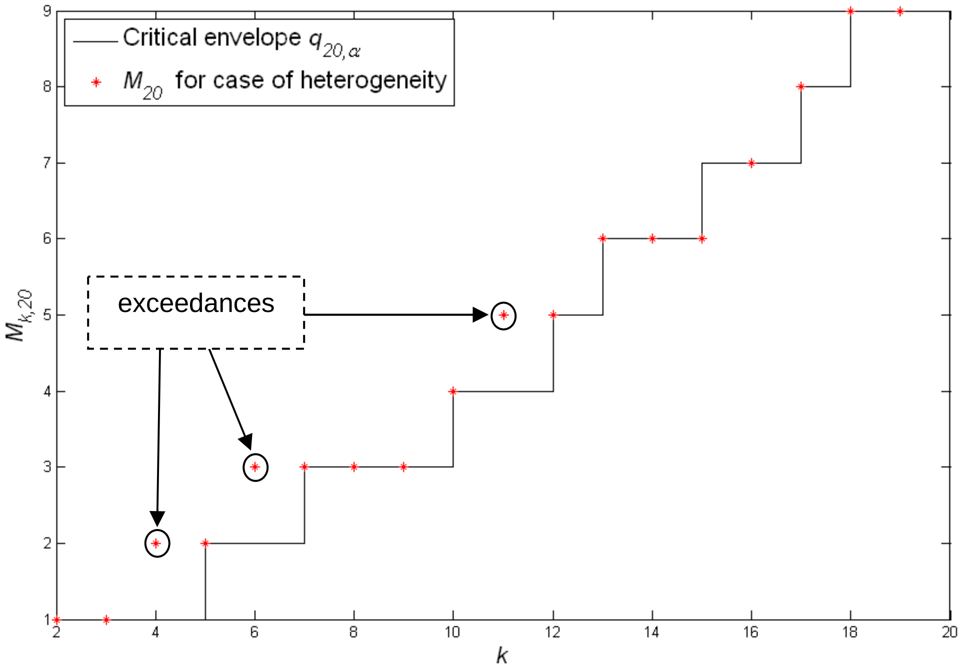

Figure 10. NAP test statistic for Figure 7 data, change point detected.

We note here at least two potential advantages of the NAP test over the SPM test. First, by design, the NAP test allows one to narrow the window of the test for possible change points, which the SPM test does not allow. In other words, if there is prior information that allows a possible change point to be restricted to a subinterval $[k_0, k_1]$ with $k_0 > 2$ and $k_1 < N$, then the critical envelope calculation is adjusted accordingly. A second advantage is that the NAP test gives information regarding not only *if* a change point exists, but also *when* it occurred. For any case where at least one exceedance exists, let $k^* = \min\{k : M_{k,N} > q_{k,N}\}$. We expect that earlier change points would be identified by smaller values of k^*. We examine the performance of the NAP test in the simulation study presented in Chapter IV.

C. THE ENSEMBLE SUM OF PAIR-MAXIMA TEST

The minimum cost assignment obtained by optimal non-bipartite weighted matching is associated with a random sample; therefore, the assignment is optimal only to the specific data at hand. Another sample with the same underlying distribution(s) would almost certainly result in a different matching with respect to sequence labels. It is natural then to examine sub-optimal (but good) matchings for additional information regarding homogeneity, and evaluate whether the information in this ensemble of matchings yields greater power to detect whether a distribution change has occurred. In particular, we consider collections of matchings that are **orthogonal**, meaning they share no common pair (this is similar to the approach of Friedman and Rafsky (1979) where they examine orthogonal minimal spanning trees). We discuss a few properties of such collections as background and then introduce a test statistic based on collections of orthogonal matchings.

1. Orthogonal Successive Optimal Matchings

We use the term **orthogonal successive optimal matchings** to refer to matchings constructed by the following process: compute an optimal non-bipartite matching on the original data, then the next best matching that is orthogonal to the first, then the next best matching that is orthogonal to the first and second, and so on. Given $N = 2n$ observations (we assume *N* even as before to simplify exposition) and some associated cost function, orthogonal successive optimal matchings have the following properties.

<u>Property 1</u>: At least $N/2$ matchings may be obtained by orthogonal successive optimal matching.

<u>Proof of Property 1</u>: The following lemma follows directly from Theorem 6.6 of Chartrand and Zhang (2005): "If a graph G has $N = 2n$ vertices and each vertex has degree at least *n*, then *G* has a perfect matching." Let $G_0 = (V, E_0)$ be the original graph on *N* vertices and $N(N-1)/2$ edges and let $G_i = (V, E_i)$ denote the subgraph whose edge set $E_i \subseteq E_0$ consists of those edges that have *not* been utilized in matchings $1, \ldots, i$.

At the beginning of the matching process, each vertex in G_0 has degree $N-1$. After the first matching $N/2$ edges are removed from G_0 with each vertex incident to exactly one such edge, so each vertex in G_1 has degree $N-2$ and a perfect matching exists on G_1 by the lemma. After $N/2-1$ orthogonal successive optimal matchings have been computed, $G_{N/2-1}$ has degree $(N-1)-(N/2-1)=N/2=n$, so at least one more matching exists by the lemma. ■

Property 2: At most $N-1$ matchings may be obtained by orthogonal successive optimal matching.

Proof of Property 2: From the discussion in Property 1, each vertex in graph G_i has degree $N-1-i$. Therefore, G_{N-1} has no edges, and no more successive matchings exist. ■

The bounds associated with Properties 1 and 2 are both strong bounds in the sense that there exist cases where no more than $N/2$ matchings may be obtained by orthogonal successive optimal matching, and also cases where exactly $N-1$ orthogonal matchings may be obtained. The following examples demonstrate each case.

Example for which no more than N / 2 matchings may be obtained: Let $N=6$, and consider a regular hexahedron under a Euclidean cost function in three dimensions (that is, a polyhedron with 6 triangular faces and all edges of equal length). Label its five vertices as shown in Figure 11.

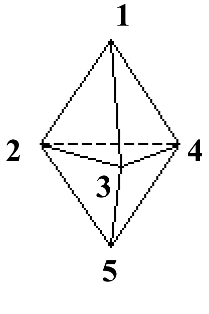

Figure 11. Example for which no more than $N/2$ matchings may be obtained.

Without loss of generality, assume each edge length of this shape equals 1. Now insert a sixth point in the center of this shape (that is, at the midpoint of the segment connecting the vertices 1 and 5) and edges to connect it to all other five vertices. The resulting cost matrix is given in Table 2:

c_{ij}	1	2	3	4	5	6
1	0	1	1	1	$2\sqrt{2/3}$	$\sqrt{2/3}$
2	1	0	1	1	1	$\sqrt{1/3}$
3	1	1	0	1	1	$\sqrt{1/3}$
4	1	1	1	0	1	$\sqrt{1/3}$
5	$2\sqrt{2/3}$	1	1	1	0	$\sqrt{2/3}$
6	$\sqrt{2/3}$	$\sqrt{1/3}$	$\sqrt{1/3}$	$\sqrt{1/3}$	$\sqrt{2/3}$	0

Table 2. Cost matrix for the regular hexahedron in Figure 11.

It is quickly verified that an optimal matching in this case pairs vertices 1, 5, and 6 with any one of vertices 2, 3, and 4 so as to make a matching. Regardless of tiebreaking procedure, the first three successive matchings exhaust all pairings of 1, 5, and 6 with 2, 3, and 4. To obtain a fourth matching, vertices 2, 3, and 4 may only be paired among each other, in which case no partner exists for the third. Therefore, no more than the $N/2$ orthogonal successive optimal matchings guaranteed by Property 1 can be constructed.

<u>Example for which exactly $N - 1$ matchings may be obtained</u>: Let $N = 4$, and consider a square under a Euclidean cost function with its vertices labeled as shown in Figure 12:

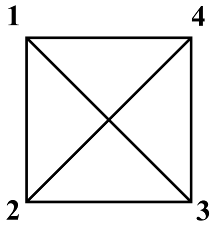

Figure 12. Example for which exactly *N* – 1 matchings may be obtained.

By inspection it is clear that, in order, $\{\{1,2\},\{3,4\}\}$, $\{\{1,4\},\{2,3\}\}$, $\{\{1,3\},\{2,4\}\}$ is an example of $N-1=3$ orthogonal successive optimal matchings.

In fact, the set of all possible collections of $N-1$ orthogonal successive optimal matchings is isomorphic to the set of all symmetric Latin squares of order N with the property that the integer N is on the diagonal. Recall that a Latin square of order N is an array consisting of the integers $\{1,\ldots,N\}$ such that each integer occurs exactly once in each row and once in each column. The isomorphism may be described as follows. Denote the $N-1$ orthogonal successive optimal matchings by $\{M_1,\ldots,M_{N-1}\}$, where $M_j=\{(u_{1j},y_{1j}),\ldots,(u_{nj},y_{nj})\}$ is the *j*th matching and u_{ij} and y_{ij} are the minimum and maximum sequence labels, respectively, of the *i*th pair listed in the *j*th matching. Enter the integer j in an $N\times N$ array at entries (u_{ij},y_{ij}) and (y_{ij},u_{ij}) for all $i\in\{1,\ldots,n\}$ and all $j\in\{1,\ldots,N-1\}$, and enter *N* on the array diagonal. The resulting Latin square carries the information indicating the stage in the succession of matchings at which the observation whose sequence number corresponds to row (column) *a* is paired with the observation whose sequence number corresponds to column (row) *b*, $a\neq b$. We illustrate the isomorphism with a very simple example on $N=6$ observations in Figure 13.

5 successive matchings

$M_1=\{\{1,2\},\{3,5\},\{4,6\}\}$

$M_2=\{\{1,3\},\{2,6\},\{4,5\}\}$

$M_3=\{\{1,4\},\{2,3\},\{5,6\}\}$

$M_4=\{\{1,5\},\{2,4\},\{3,6\}\}$

$M_5=\{\{1,6\},\{2,5\},\{3,4\}\}$

6×6 Latin square

6	1	2	3	4	5
1	6	3	4	5	2
2	3	6	5	1	4
3	4	5	6	2	1
4	5	1	2	6	3
5	2	4	1	3	6

Figure 13. Five orthogonal successive optimal matchings on $N=6$ observations with associated Latin square. Cost function does not necessarily satisfy the triangle inequality.

In our research, most random samples we observed of size N admitted $N-1$ orthogonal successive optimal matchings. However, the fact that this is not always the case requires that our theory accommodate cases where less than $N-1$ orthogonal successive optimal matchings are admitted.

2. The K_N Statistic

We proceed to formulate a test statistic based on orthogonal successive optimal matchings. Let $T_{N,i}$ denote the sum of pair-maxima statistic associated with the i^{th} best orthogonal matching. It is straightforward to show that $T_{N,i}$ is marginally distributed as T_N. We prove this for the case of continuous random variables.

<u>Proof</u>: Let $\left(X_{(1)},\ldots,X_{(N)}\right)$ be some standard ordering of observations $\left(X_1,\ldots,X_N\right)$ based solely on data content (e.g., an ordering based on observation norm) and let C be the cost matrix with respect to this standard ordering. Let $\mathcal{V}$ denote the set of all $N\times N$ 0-1 matrices associated with perfect matchings on N observations; that is, $\mathcal{V}$ is the set of all symmetric matrices that have a "1" entry for pairings in a matching and "0" otherwise. Two matchings $U,V\in\mathcal{V}$ are orthogonal if and only if $U\circ V=\mathbf{0}$, where "$\circ$" is the Hadamard (coordinate-wise) product of U and V. Then the optimal non-bipartite matching problem can be expressed as

$$\begin{aligned}&\min\ \operatorname{tr}\left(CV\right)\\&\text{subject to } V\in\mathcal{V},\end{aligned}\tag{3.59}$$

where $\operatorname{tr}\left(A\right)$ denotes the trace of matrix A. Now let V_1^* be the solution to (3.59), and define nested sets $\mathcal{V}_1\equiv\mathcal{V}\supset\mathcal{V}_2\supset\cdots\supset\mathcal{V}_i$, orthogonal solutions $V_2^*,V_3^*,\ldots,V_i^*$, and objective values $z_1^*,z_2^*,\ldots,z_i^*$ for $i\in\left\{1,\ldots,N/2\right\}$ recursively as follows:

$$
\begin{aligned}
&V_1^* \text{ solves } z_1^* = \min \operatorname{tr}(CV) \\
&\qquad\qquad\qquad \text{subject to } V \in \mathcal{V}_1, \\
&V_2^* \text{ solves } z_2^* = \min \operatorname{tr}(CV) \\
&\qquad\qquad\qquad \text{subject to } V \in \mathcal{V}_2 = \left\{ V \in \mathcal{V} : V \circ V_1^* = \mathbf{0} \right\}, \\
&\qquad\qquad\quad \vdots \\
&V_i^* \text{ solves } z_i^* = \min \operatorname{tr}(CV), \\
&\qquad\qquad\qquad \text{subject to } V \in \mathcal{V}_i = \left\{ V \in \mathcal{V} : V \circ V_j^* = \mathbf{0} \ \ \forall\, j \in \{1,2,\ldots,i-1\} \right\}.
\end{aligned}
\tag{3.60}
$$

For continuous random variables, $z_1^* < z_2^* < \cdots < z_i^*$ with probability 1.

Now, for any permutation π applied to the integers $\{1,\ldots,N\}$, let A_π denote the associated $N \times N$ permutation matrix. If the order of $\left(X_{(1)},\ldots,X_{(N)}\right)$ is permuted by π, then the cost matrix for the permuted observations is $A_\pi C A_\pi'$. Define nested sets $\mathcal{V}_{\pi,1} \equiv \mathcal{V} \supset \mathcal{V}_{\pi,2} \supset \cdots \supset \mathcal{V}_{\pi,i}$, orthogonal solutions $V_{\pi,2}^*, V_{\pi,3}^*, \ldots, V_{\pi,i}^*$, and objective values $z_{\pi,1}^*, z_{\pi,2}^*, \ldots, z_{\pi,i}^*$ for $i \in \{1,\ldots,N/2\}$ recursively like before:

$$
\begin{aligned}
&V_{\pi,1}^* \text{ solves } z_{\pi,1}^* = \min \operatorname{tr}(C_\pi V) \\
&\qquad\qquad\qquad \text{subject to } V \in \mathcal{V}_{\pi,1}, \\
&\qquad\qquad\quad \vdots \\
&V_{\pi,i}^* \text{ solves } z_{\pi,i}^* = \min \operatorname{tr}(C_\pi V), \\
&\qquad\qquad\qquad \text{subject to } V \in \mathcal{V}_{\pi,i} = \left\{ V \in \mathcal{V} : V \circ V_{\pi,j}^* = \mathbf{0} \ \ \forall j \in \{1,2,\ldots,i-1\} \right\}.
\end{aligned}
\tag{3.61}
$$

Note that $\operatorname{tr}(C_\pi V) = \operatorname{tr}(A_\pi C A_\pi' V) = \operatorname{tr}(C A_\pi' V A_\pi)$ for all $V \in \mathcal{V}$ by the permutation invariance of the trace operator.

We now show by induction that for any *i*, $V_{\pi,i}^* = A_\pi V_i^* A_\pi'$. The assertion is true for $i = 1$, since $A_\pi V_1^* A_\pi' \in \mathcal{V} = \mathcal{V}_{\pi,1}$,

$$
z_{\pi,1}^* = \min_{V \in \mathcal{V}_{\pi,1}} \operatorname{tr}(C_\pi V) = \min_{V \in \mathcal{V}} \operatorname{tr}(C A_\pi' V A_\pi) = \min_{U \in \mathcal{V}} \operatorname{tr}(CU) = z_1^*,
\tag{3.62}
$$

and

$$
\operatorname{tr}\left(C_\pi \left(A_\pi V_1^* A_\pi'\right)\right) = \operatorname{tr}\left(C V_1^*\right) = z_1^*.
\tag{3.63}
$$

Now suppose $V^*_{\pi,j} = A_\pi V^*_j A'_\pi$ for all $j \in \{1,\ldots,k\}$, $k < N/2$. Then

$$\begin{aligned} \mathcal{V}_{\pi,k+1} &= \left\{ V \in \mathcal{V} : V \circ V^*_{\pi,j} = \mathbf{0} \quad \forall\, j \in \{1,2,\ldots,k\} \right\} \\ &= \left\{ V \in \mathcal{V} : A'_\pi \left(V \circ V^*_{\pi,j} \right) A_\pi = \mathbf{0} \quad \forall\, j \in \{1,2,\ldots,k\} \right\} \\ &= \left\{ V \in \mathcal{V} : \left(A'_\pi V A_\pi \right) \circ \left(A'_\pi V^*_{\pi,j} A_\pi \right) = \mathbf{0} \quad \forall\, j \in \{1,2,\ldots,k\} \right\} \\ &= \left\{ V \in \mathcal{V} : A'_\pi V A_\pi \circ V^*_j = \mathbf{0} \quad \forall\, j \in \{1,2,\ldots,k\} \right\}. \end{aligned} \tag{3.64}$$

Therefore, the problem

$$\begin{gathered} \min \operatorname{tr}\left(C_\pi V \right) \\ \text{subject to } V \in \mathcal{V}_{\pi,k+1} \end{gathered} \tag{3.65}$$

has the same solution as

$$\begin{gathered} \min \operatorname{tr}\left(C A'_\pi V A_\pi \right) \\ \text{subject to } A'_\pi V A_\pi \in \mathcal{V}_{k+1}\ , \end{gathered} \tag{3.66}$$

and the solution to (3.66) is $V = A_\pi V^*_{k+1} A'_\pi$. Therefore, $V^*_{\pi,k+1} = A_\pi V^*_{k+1} A'_\pi$, and thus by induction $V^*_{\pi,i} = A_\pi V^*_i A'_\pi$ for any *i*.

Finally, let ρ be the antirank vector for the ordered observations so that $X_{(j)} = X_{\rho(j)}$. Then for any *i*, if V^*_i solves the i^{th} orthogonal matching problem with respect to the standard ordering, then $V^*_{\rho,i}$ is the solution with respect to the original ordering. Conditional on $\left(X_{(1)}, \ldots, X_{(N)} \right)$, ρ is uniformly distributed over all $N!$ possible permutations applied to the integers $\{1,\ldots,N\}$ under the null hypothesis. So, each element of $\mathcal{V}$ is equally likely to be the i^{th} orthogonal successive optimal matching, and thus each possible labeling of vertices in that matching is equally likely. Therefore, $T_{N,i}$ is marginally distributed as T_N. ■

By Property 1 above, $T_{N,i}$ is well-defined for all $i \le N/2$. We define $S_{N,k}$ to be the cumulative sum of pair-maxima over the first k matchings:

$$S_{N,k} = \sum_{i=1}^{k} T_{N,i}\ . \tag{3.67}$$

Likewise, $S_{N,k}$ is well-defined for all $k \le N/2$. The mean of $S_{N,k}$ under the null hypothesis is computed directly:

(3.68) $$\mu_{N,k} = E\left[S_{N,k}\right] = \sum_{i=1}^{k} E\left[T_{N,i}\right] = k\mu_N = k\frac{N(N+1)}{3}.$$

Just as T_N takes on smaller values under alternative hypotheses than under the null hypothesis, we expect that under alternative hypotheses $S_{N,k}$ will tend to deviate below its mean value. Therefore, we define

(3.69) $$K_N = \max_{k \in \{1,2,\ldots,N/2\}} \left(\frac{\mu_{N,k} - S_{N,k}}{c_N} \right),$$

to be our *Ensemble Sum of Pair-Maxima* (ESPM) test statistic, where

(3.70) $$c_N = (N-1)\sqrt{\frac{N(N+1)}{180}}$$

is a scaling constant whose choice is motivated in the following section. So, K_N measures the maximum cumulative deviation of the $T_{N,i}$ from their mean over $N/2$ orthogonal successive optimal matchings. For convenience we choose K_N to be the negative of the typical centered random variable so that smaller values of $S_{N,k}$ (which are evidence of an underlying distribution change) correspond to larger values of K_N.

3. Brownian Bridge Motivation for K_N

The formulation of the K_N statistic is based in part on structural similarities of the sequence $\left(S_{N,1},\ldots,S_{N,N/2}\right)$ to a Brownian bridge. Recall that a stochastic process $\{W(t), t \ge 0\}$ is called a **Gaussian process** if $\left(W(t_1),\ldots,W(t_j)\right)$ has a multivariate normal distribution for all $(t_1,\ldots,t_j)$ and for all $j \in \{1,2,\ldots\}$, and that a Gaussian process $\{B(t), 0 \le t \le 1\}$ with $E\left[B(t)\right] = 0,\ 0 \le t \le 1$, and $\operatorname{Cov}\left(B(s),B(t)\right) = s(1-t)$, $0 \le s \le t \le 1$, is called a **Brownian bridge** (Ross, 2003, pp. 622-623).

We desire an expression for the variance of $S_{N,k}$ to compare the sequence $\left(S_{N,1},\ldots,S_{N,N/2}\right)$ to a Brownian bridge. Such an expression depends on the covariance between the $T_{N,i}$, which is difficult to determine analytically. However, simulation suggests that $\mathrm{Cov}\left(T_{N,i},T_{N,j}\right)=\mathrm{Cov}\left(T_{N,1},T_{N,2}\right)$ for all $i \neq j$, so for the remainder of this section we *assume* this to be true for the sake of comparison to a Brownian bridge only.

Under the assumption that $\mathrm{Cov}\left(T_{N,i},T_{N,j}\right)=\mathrm{Cov}\left(T_{N,1},T_{N,2}\right)$ for all $i \neq j$, it follows that

$$\begin{aligned}\underset{\sim}{\sigma}^2_{N,k} &= Var\left[S_{N,k}\right]=k\sigma_N^2+2\sum_{i=2}^{k}\sum_{j=1}^{i-1}\mathrm{Cov}\left(T_{N,i},T_{N,j}\right)\\ &= k\sigma_N^2+k\left(k-1\right)\mathrm{Cov}\left(T_{N,1},T_{N,2}\right)\ \forall\, k,\end{aligned}\tag{3.71}$$

where the underscore-tilde notation "$\underset{\sim}{\sigma}^2_{N,k}$" indicates that equality depends on our covariance assumption. It is straightforward to solve for $\mathrm{Cov}\left(T_{N,1},T_{N,2}\right)$ under this assumption by observing that in any case for which $N-1$ orthogonal successive optimal matchings can be constructed, every possible pairing of observations has been considered. For this case then,

$$S_{N,N-1}=\left(N-1\right)\mu_{N,N-1},\tag{3.72}$$

which is constant for fixed *N*. Therefore $Var\left[S_{N,N-1}\right]=0$. Applying this boundary-value condition to (3.71) gives

$$0=Var\left[S_{N,N-1}\right]=\left(N-1\right)\sigma_N^2+\left(N-1\right)\left(N-2\right)\mathrm{Cov}\left(T_{N,1},T_{N,2}\right)\ \forall\, N,\tag{3.73}$$

so

$$\mathrm{Cov}\left(T_{N,1},T_{N,2}\right)=-\frac{\sigma_N^2}{\left(N-2\right)}=-\frac{N\left(N-2\right)\left(N+1\right)}{2\left(N-2\right)90}=-\frac{N\left(N+1\right)}{180}\ \forall\, N.\tag{3.74}$$

Therefore, we obtain the desired result

$$\begin{aligned} \underset{\sim}{\sigma}^2_{N,k} &= k\sigma_N^2 + k(k-1)\operatorname{Cov}(T_{N,1}, T_{N,2}) \\ &= k\sigma_N^2 - k(k-1)\frac{N(N+1)}{180} \\ &= \frac{kN(N+1)(N-k-1)}{180}. \end{aligned} \tag{3.75}$$

Finally, define $S_{N,0} \equiv 0$ and $\mu_{N,0} \equiv 0$, let $t = k/(N-1)$ for $k \in \{1,\dots,N/2\}$, and define stochastic process $B_N(t)$ by

$$B_N(t) \equiv \frac{\mu_{N,t(N-1)} - S_{N,t(N-1)}}{c_N},\ t \in \left\{0, \frac{1}{N-1}, \frac{2}{N-1}, \dots, \frac{N/2}{(N-1)}\right\},\ \ \forall N. \tag{3.76}$$

Then for all s and t such that $s \le t$ and $s,t \in \{0, 1/(N-1), \dots, N/(2N-2)\}$ and for all even *N* we have

$$B_N(0) = 0, \tag{3.77}$$

$$E[B_N(t)] = 0, \tag{3.78}$$

and

$$
\begin{aligned}
\operatorname{Cov}\left[B_N(s), B_N(t)\right] &= \operatorname{Cov}\left[\frac{\mu_{N,s(N-1)} - S_{N,s(N-1)}}{c_N}, \frac{\mu_{N,t(N-1)} - S_{N,t(N-1)}}{c_N}\right] \\
&= \left(1/c_N^2\right)\operatorname{Cov}\left[S_{N,s(N-1)}, S_{N,t(N-1)}\right] \\
&= \left(1/c_N^2\right)\operatorname{Cov}\left[S_{N,s(N-1)}, S_{N,s(N-1)} + \sum_{j=s(N-1)+1}^{t(N-1)} T_{N,j}\right] \\
&= \left(1/c_N^2\right)\left(\operatorname{Var}\left[S_{N,s(N-1)}\right] + \operatorname{Cov}\left[\sum_{i=1}^{s(N-1)} T_{N,i}, \sum_{j=s(N-1)+1}^{t(N-1)} T_{N,j}\right]\right) \\
&= \left(1/c_N^2\right)\left(\operatorname{Var}\left[S_{N,s(N-1)}\right] + \sum_{i=1}^{s(N-1)} \sum_{j=s(N-1)+1}^{t(N-1)} \operatorname{Cov}\left[T_{N,i}, T_{N,j}\right]\right) \\
&= \left(1/c_N^2\right)\left(\underset{\sim}{\sigma}^2_{N,s(N-1)} + \sum_{i=1}^{s(N-1)} \sum_{j=s(N-1)+1}^{t(N-1)} \operatorname{Cov}\left[T_{N,1}, T_{N,2}\right]\right) \\
&= s(1-s) + \left(1/c_N^2\right) s(N-1)(t-s)(N-1)\left(-\frac{N(N+1)}{180}\right) \\
&= s(1-s) - s(t-s) \\
&= s(1-t). \qquad (3.79)
\end{aligned}
$$

This structure of $\left\{B_N(t), t \in \left\{0, 1/(N-1), \ldots, N/(2N-2)\right\}\right\}$ for our choice of c_N and equal covariance assumption suggests a connection to the Brownian bridge.

Shorack and Wellner (1986) present several useful results pertaining to the Brownian bridge, including the following:

If $\left\{B(t), 0 \le t \le 1\right\}$ *is a Brownian bridge, then for all* $a, b > 0$

$$
\begin{aligned}
&P\left(B(t) \le a(1-t) + bt \text{ for } 0 \le s \le t \le u \le 1 \,\middle|\, B(s) = x \text{ and } B(u) = y\right) \\
&\qquad = 1 - \exp\left\{-2\left[a(1-s) + bs - x\right]\left[a(1-u) + bu - y\right]/(u-s)\right\}.
\end{aligned} \qquad (3.80)
$$

Setting $a = b = \lambda > 0$ and $\gamma = x = 0$,

$$
\begin{aligned}
&P\left(B(t) \le \lambda \text{ for } 0 \le t \le u \le 1 \,\middle|\, B(u) = y\right) \\
&\qquad = 1 - \exp\left\{-2\lambda(\lambda - y)/u\right\}, \qquad \text{where } B(u) \sim N\left(0, u(1-u)\right).
\end{aligned} \qquad (3.81)
$$

Taking the expected value over $B(u)$ yields the identity

$$
\begin{aligned}
&P\left(B(t)\leq\lambda \text{ for } 0\leq t\leq u\leq 1,\lambda>0\right)\\
&=\frac{1}{\sqrt{2\pi u(1-u)}}\int_{-\infty}^{\lambda}\left(1-\exp\left\{-\frac{2\lambda(\lambda-y)}{u}\right\}\right)\exp\left\{-\frac{y^2}{2u(1-u)}\right\}dy\\
&=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\frac{\lambda}{\sqrt{u(1-u)}}}\left(1-\exp\left\{-\frac{2\lambda\left(\lambda-x\sqrt{u(1-u)}\right)}{u}\right\}\right)\exp\left\{-\frac{x^2}{2}\right\}dx\\
&=\Phi\left(\frac{\lambda}{\sqrt{u(1-u)}}\right)-\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\frac{\lambda}{\sqrt{u(1-u)}}}\exp\left\{-\frac{1}{2}\left(x-\frac{2\lambda\sqrt{u(1-u)}}{u}\right)^2\right\}\exp\left\{-2\lambda^2\right\}dx\\
&=\Phi\left(\frac{\lambda}{\sqrt{u(1-u)}}\right)-\exp\left\{-2\lambda^2\right\}\Phi\left(\frac{\lambda(2u-1)}{\sqrt{u(1-u)}}\right),
\end{aligned}
\tag{3.82}
$$

where Φ is the standard normal cumulative distribution function. For $u=1/2$, equation (3.82) reduces to

$$P\left(B(t)\leq\lambda \text{ for } 0\leq t\leq 1/2,\lambda>0\right)=\Phi(2\lambda)-\frac{1}{2}\exp\left\{-2\lambda^2\right\}. \tag{3.83}$$

Now define $K=\sup_{t\in[0,1/2]}B(t)$ for Brownian bridge $\left\{B(t),0\leq t\leq 1\right\}$. It follows directly from (3.83) that the critical value K_α corresponding to $P(K>K_\alpha)=\alpha$ is a solution to

$$1-\alpha-\Phi(2x)+\frac{1}{2}\exp\left\{-2x^2\right\}=0, \tag{3.84}$$

which can be well-approximated using standard computing software. In other words, the null distribution of *K* is known and so *K* is an obvious choice for a test statistic *if* a process of interest is a Brownian bridge. The question at this point then is, "Does the process $\left\{B_N(t),t\in\left\{0,1/(N-1),\ldots,N/(2N-2)\right\}\right\}$ asymptotically approach a Brownian bridge?" Under our covariance assumption, this process has the same mean and covariance structure as a Brownian bridge for all *N*. Furthermore $B_N(t)$ is by construction a shifted, scaled sum of random variables whose marginal distributions are asymptotically normal. Nevertheless, we are somewhat surprised to observe that $B_N(t)$ is *not* normal even for fairly large *N*. Figures 14-18 show normal quantile-quantile plots of $B_N(t)$ for 10,000 simulations associated with the null hypothesis, using a standard bivariate normal distribution, $N = 300$, and $t = k/(N-1)$ for $k = 1, 10, 30, 100, \text{and } 150$. Figure 14 shows $k = 1$, which of course is simply $(T_N - \mu_N)/c_N$. We have proven already that T_N is asymptotically normal, and indeed Figure 14 is confirming evidence.

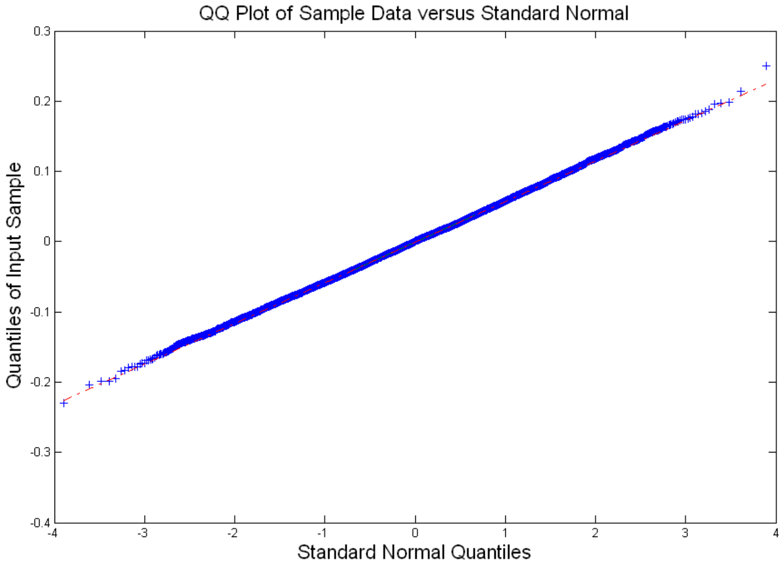

Figure 14. **Quantile-Quantile plot of $B_N(k/(N-1))$ for 10,000 simulations of N observations from a standard bivariate normal distribution; $N = 300$, $k = 1$.**

As *k* varies from 1 to $N/2$, $B_N(t)$ exhibits markedly non-normal behavior as demonstrated in panels Figures 15-18. In Figure 15, for $k = 10$ (corresponding to $t \approx 0.033$) $B_N(t)$ shows signs of slight positive skewness. This skewness is much more apparent as *k* approaches the middle of the interval as seen in Figures 16-18 for $k = 30$

($t \approx 0.10$), $k = 100$ ($t \approx 0.33$), and $k = 150$ ($t \approx 0.50$). In other words, $B_N(t)$ constitutes an unusual "natural" example of a case where the distribution of the sum of a large number of asymptotically normal random variables does *not* appear to approach a normal distribution.

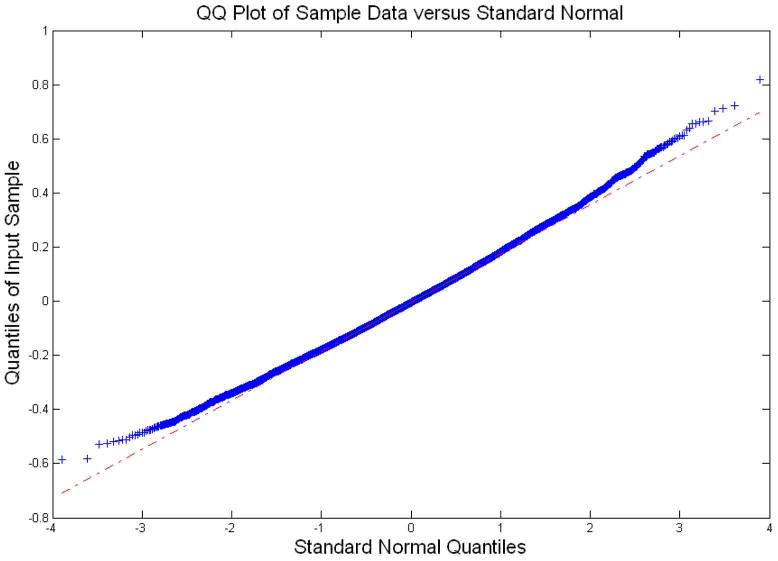

Figure 15. Quantile-Quantile plot of $B_N\left(k/(N-1)\right)$ for 10,000 simulations of N observations from a standard bivariate normal distribution; $N = 300$, $k = 10$.

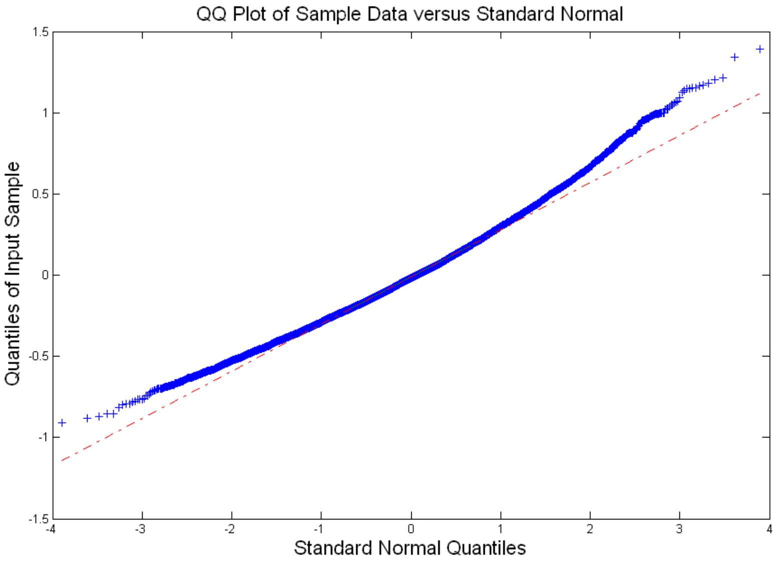

Figure 16. **Quantile-Quantile plot of $B_N\left(k/\left(N-1\right)\right)$ for 10,000 simulations of N observations from a standard bivariate normal distribution; $N=300$, $k=30$.**

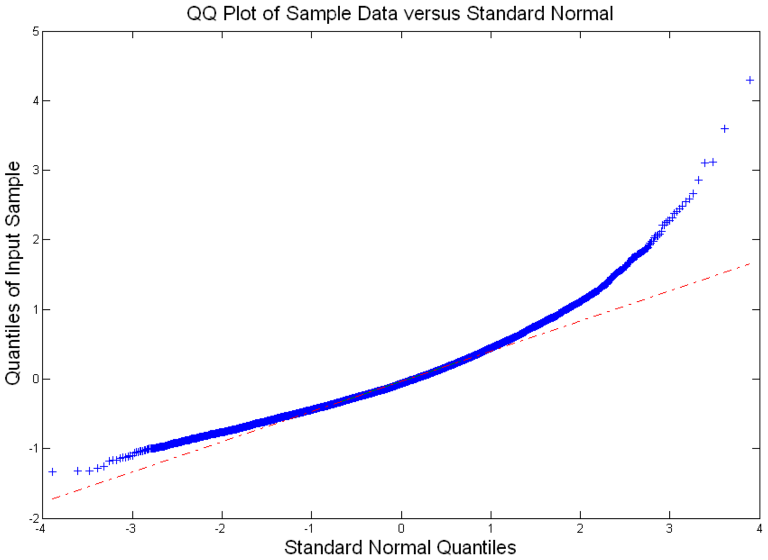

Figure 17. **Quantile-Quantile plot of $B_N\left(k/\left(N-1\right)\right)$ for 10,000 simulations of N observations from a standard bivariate normal distribution; $N=300$, $k=100$.**

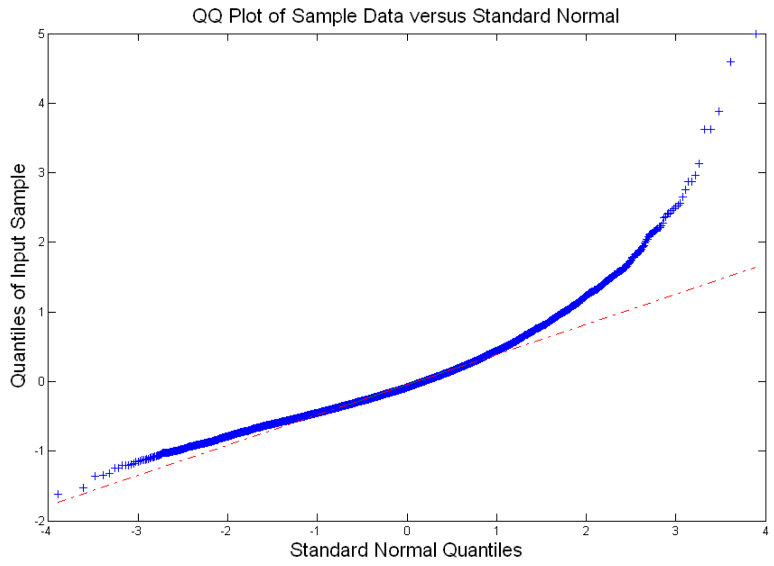

Figure 18. Quantile-Quantile plot of $B_N\left(k/(N-1)\right)$ for 10,000 simulations of N observations from a standard bivariate normal distribution; $N = 300$, $k = 150$.

Of course, this evidence alone does not prove that $B_N(t)$ fails to approach a Brownian bridge in the limit; however, it does establish that even for fairly large N a Brownian bridge approximation may not be a good one. Indeed, simulation shows that for fairly large N (from 100 to 300) a Brownian bridge approximation for $B_N(t)$ gives reasonable tail probability estimates for α near .10, but less so for smaller values of α. Table 3 shows tail probability results based on 10,000 simulations of N observations from a standard bivariate normal distribution for $\alpha = 0.10$, 0.05, 0.025, and 0.01 where the achieved test level for each combination of N and α is the fraction of simulations for which $K_N = \max_{k\in\{0,1,\ldots,N/2\}} B_N\left(k/(N-1)\right) > \kappa_\alpha$, where κ_α is a root of (3.84).

α	κ_α	Achieved test level		
		$N=100$	$N=200$	$N=300$
0.10	0.9757	0.0877	0.0980	0.0984
0.05	1.1334	0.0586	0.0649	0.0635
0.025	1.2731	0.0385	0.0432	0.0419
0.01	1.4382	0.0248	0.0278	0.0266

Table 3. Achieved test levels for $\max_{k\in\{0,1,\ldots,N/2\}} B_N\left(k/(N-1)\right)$ using Brownian bridge critical values for 10,000 simulations of N observations from a standard bivariate normal distribution. Achieved test level for each combination of N and α is the fraction of simulations for which $\max_{k\in\{0,1,\ldots,N/2\}} B_N\left(k/(N-1)\right)>\kappa_\alpha$.

Because the null distribution for K_N is difficult to obtain exactly, we obtain useful tail probability approximations by simulation. These tail probability approximations depend on both sample size N and dimensionality d. Approximate critical values for K_N based on simulation are provided in Appendix B for various values of N, d, and α. As indicated in the SPM test discussion, the ESPM test proves to be significantly more powerful than a single SPM test. We show performance results in the next chapter.

We close this discussion of an SPM-based ensemble statistic with the comment that it is less clear exactly how to formulate an analogous ensemble extension for the NAP statistic. In that case, it seems natural to find the sequence of orthogonal best matchings as in the ESPM case and then compute $\mathbf{M}_N$ for each matching. Letting $\mathbf{M}_{N,i}$ denote the $\mathbf{M}_N$ statistic associated with the i^{th} best orthogonal matching, one would expect to be able to extract more change-point information out of the collection of vectors $\left\{\mathbf{M}_{N,1},\mathbf{M}_{N,2},\ldots,\mathbf{M}_{N,N-1}\right\}$ than out of the NAP test using $\mathbf{M}_{N,1}$ alone. For now, we leave study of an ensemble NAP statistic for future work.

D. THE BIPARTITE ACCUMULATED PAIRS TEST

The tests introduced thus far are based on non-bipartite weighted matchings, and they test for a change point over an entire set of observations for which there is no control set. That is, no prior information exists regarding the in-control distribution. In many cases, however, some history of observations which are known (or assumed) to be in-control is available, and the problem is to determine whether a change point exists in a set of future observations. Therefore, we propose the *Bipartite Accumulated Pairs* (BAP) test for cases where in-control observation history is available. As the name indicates, the BAP test is constructed using *bipartite* matchings (as opposed to non-bipartite matchings as in all our previous tests). Recall that a bipartite matching pairs observations from one pre-designated group with observations from another, and is a solution to the integer program (2.13).

1. The Z Statistic

Assume we have some history $\{X_1,\ldots,X_m\}$ of *m* control observations, and we desire to test whether a change point exists in a sequence $(X_{m+1},\ldots,X_{m+n})$ of *n* new test observations. One approach to this problem is to estimate the in-control distribution based on the observation history and then test whether it is likely that the new observations are drawn from the estimated distribution. An alternative matching-based approach is to compute an optimal bipartite matching between the control and test observations and use the information in the matching to test whether a change point exists in the test data.

We employ the following approach for the case $m < n$ (more test data than control data; we discuss the $m \geq n$ case later): Compute a minimum-weight bipartite matching which pairs each control observation with some test observation, based on some predetermined cost function. We emphasize here that some test observations will necessarily remain unpaired, unlike the non-bipartite matching case where at most one observation is left unpaired. Define random variables $Z_1,\ldots,Z_{n-1}$ where Z_k is the number of test observations among $\left\{X_{m+1},\ldots,X_{m+k}\right\}$ which are paired with control observations. Notice that this test has very much the same flavor as the NAP test, as it counts the accumulation of pairs in the optimal matching with respect to the order of the test data. By construction the Z_k are dependent random variables each of which is marginally distributed as hypergeometric. In fact, $Z_k = Z_{k-1} + \delta_k$, where $\delta_k = 1$ if test observation Z_{m+k} is paired with a control observation and $\delta_k = 0$ otherwise. So,

$$(3.85) \qquad P\left(Z_k = r\right) = \frac{\binom{m}{r}\binom{n-m}{k-r}}{\binom{n}{k}}, \quad r = \max\left(0, m+k-n\right),\ldots,\min\left(m,k\right).$$

This test is designed to test null hypothesis $H_0 : F = G$ against alternative $H_1 : F \neq G$ where $X_1,\ldots,X_m \sim F$, $X_{m+1},\ldots,X_{m+n} \sim G$, and F and G are unknown. If a change point is present in the sequence of test observations, one expects that more pairs will be formed between the first k test observations and the control observations than if no change point is present, so the existence of a change point should be seen in pairings being "front-loaded" in the Z_k sequence. We call the vector $\mathbf{Z} = \left(Z_1,\ldots,Z_{n-1}\right)'$ the Bipartite Accumulated Pairs (BAP) test statistic.

2. Critical Envelope

We develop an exact simultaneous test for $\mathbf{Z}$ in a manner closely following that for $\mathbf{M}_N$ of the Non-Bipartite Accumulated Pairs (NAP) test. Specifically, we seek a vector of non-negative integers $\mathbf{q}_\alpha = \left\{q_1,\ldots,q_{n-1}\right\}$ so that the following is true for a given test level α:

$$(3.86) \qquad P\left(Z_k \le q_k,\ 1 \le k \le n-1\right) \ge 1-\alpha .$$

This allows us to construct a simultaneous level α test. As for the NAP case, we take q_k to be the $1-\tilde{\alpha}$ quantile of the distribution of Z_k (hypergeometric) so that the non-simultaneous test at stage k has level $\tilde{\alpha}$; that is, $P\left(Z_k > q_k\right) \le \tilde{\alpha}$ for each $k \in \{1,\ldots,n-1\}$. Then we use a recursive computational scheme to select $\tilde{\alpha}$ that gives a simultaneous test level as close to α as possible without exceeding this value. To develop the recursion, first note that the $\tilde{\alpha}$-quantiles satisfy the properties $q_k = \min\{r : P(Z_k \le r) > \tilde{\alpha}\}$ and $q_1 \le q_2 \le \cdots \le q_{n-1}$. Now using the same conditioning approach as in (3.52)-(3.58), we have

$$P\left(Z_j \le q_j, 1 \le j \le k\right) = \sum_{r=0}^{q_k} \pi\left(r,k\right) \cdot P\left(Z_k = r\right) \tag{3.87}$$

where $\pi\left(r,k\right) = P\left(Z_1 \le q_1, \ldots, Z_{k-1} \le q_{k-1} \mid Z_k = r\right)$ and $P\left(Z_k = r\right)$ is given by (3.85). Observe that $\pi\left(r,k\right)$ observes a simple recursive relationship expressed by the following lemma:

Lemma 3-6:

$$\begin{aligned} &\pi\left(r,k\right) = \frac{r}{k}\pi\left(r-1,k-1\right) I\left(r-1 \le q_{k-1}\right) + \frac{k-r}{k}\pi\left(r,k-1\right) I\left(r \le q_{k-1}\right), \\ &k = 2,\ldots,\max\left\{i : q_i < \min\left(i,m\right)\right\};\ r = 0,\ldots,q_k . \end{aligned} \tag{3.88}$$

Proof of Lemma 3-6:

$$\pi\left(r,k\right) = \sum P\left(Z_j \le q_j, 1 \le j \le k-2 \middle| Z_{k-1} = s, Z_k = r\right) \cdot P\left(Z_{k-1} \le s \middle| Z_k = r\right)$$

$$= \sum P\left(Z_j \le q_j, 1 \le j \le k-2 \middle| Z_{k-1} = s\right) \cdot P\left(Z_{k-1} \le s \middle| Z_k = r\right)$$

$$= \sum \pi\left(s,k-1\right) \cdot P\left(Z_{k-1} \le s \middle| Z_k = r\right) \cdot I\left(s \le q_{k-1}\right) \tag{3.89}$$

$$\begin{aligned} &= \pi\left(r-1,k-1\right) \cdot P\left(Z_{k-1} = r-1 \middle| Z_k = r\right) \cdot I\left(r-1 \le q_{k-1}\right) \\ &\quad + \pi\left(r,k-1\right) \cdot P\left(Z_{k-1} = r \middle| Z_k = r\right) \cdot I\left(r \le q_{k-1}\right) \end{aligned}$$

Under the null hypothesis, $Z_k = r$ implies that each of the k observations $X_{m+1}, \ldots, X_{m+k}$ is equally likely to be among r pairs in the bipartite matching; therefore $P\left(Z_{k-1} = r-1 \middle| Z_k = r\right) = \frac{r}{k}$ and $P\left(Z_{k-1} = r \middle| Z_k = r\right) = \frac{k-r}{k}$. ■

The recursion starts at $k = 2$ and continues to $k = \max\left\{i : q_i < \min\left(i, m\right)\right\}$. For $k = 2$ we have

(3.90)
$$\pi\left(r, 2\right) = \begin{cases} 1 - \left(1 - q_1\right)\dfrac{r}{2}, & 0 \le r \le q_2 \ ; \\ 0, & r > q_2 \ . \end{cases}$$

This recursion scheme is quite readily implemented in S-PLUS®, R, MATLAB®, or other interpreted languages; an implementation for R is included in Appendix C.

The simultaneous test presented here is clearly an improvement over the Bonferroni method. For example, with $m = 25$ control points and $n = 50$ test points a nominal $\alpha = 0.05$ simultaneous test achieves test level .047 using uses $\tilde{\alpha} = .011$. By contrast the Bonferroni method achieves test level .006 using $\tilde{\alpha} = .05 / 49 = .001$. The improvement is even more dramatic in larger samples. For $m = 500, n = 1000$, the BAP test achieves level .049 using $\tilde{\alpha} = .00203$ compared to an achieved level of .0018 using $\tilde{\alpha} = .05 / 999 = .00005$ by Bonferroni.

3. A Graphical Example

Figure 19 shows 20 observations all drawn from a standard bivariate normal distribution. The plot marker for each point is its sequence label. The first $m = 4$ points are the control set and are circled for emphasis; the last $n = 16$ points are the test set. Since all the test data are from the same distribution as the control data, there is no change point. We use a MATLAB® implementation of the Jonker-Volgenant assignment algorithm (1987) provided by Levedahl (2000) to compute an optimal bipartite match with respect to Euclidean distance for these data; the resulting pairs are shown connected by line segments. Table 4 shows the critical envelope $\mathbf{q}_\alpha = \left(q_1, \ldots, q_{15}\right)'$ and the test statistic $\mathbf{Z} = \left(Z_1, \ldots, Z_{15}\right)'$ for the data in Figure 19. Recall that $k = 1$ corresponds to $X_{m+1} = X_5$, $k = 2$ corresponds to $X_{m+2} = X_6$, and so on, and Z_k is equal to the number of pairings of between the sets $\left\{X_1, \ldots, X_4\right\}$ and $\left\{X_5, \ldots, X_{5+k-1}\right\}$. For this small data set we can determine the Z_k values by inspection from Figure 19. Since

X_5, X_7, X_{10}, and X_{15} are the test set elements paired with elements in the control set, $Z_1 = Z_2 = 1$, $Z_3 = Z_4 = Z_5 = 2$, $Z_6 = \cdots = Z_{10} = 3$, and $Z_{11} = \cdots = Z_{15} = 4$. None of these values exceeds the critical envelope, so we do not reject the null hypothesis that all the test data share the same distribution as the control data.

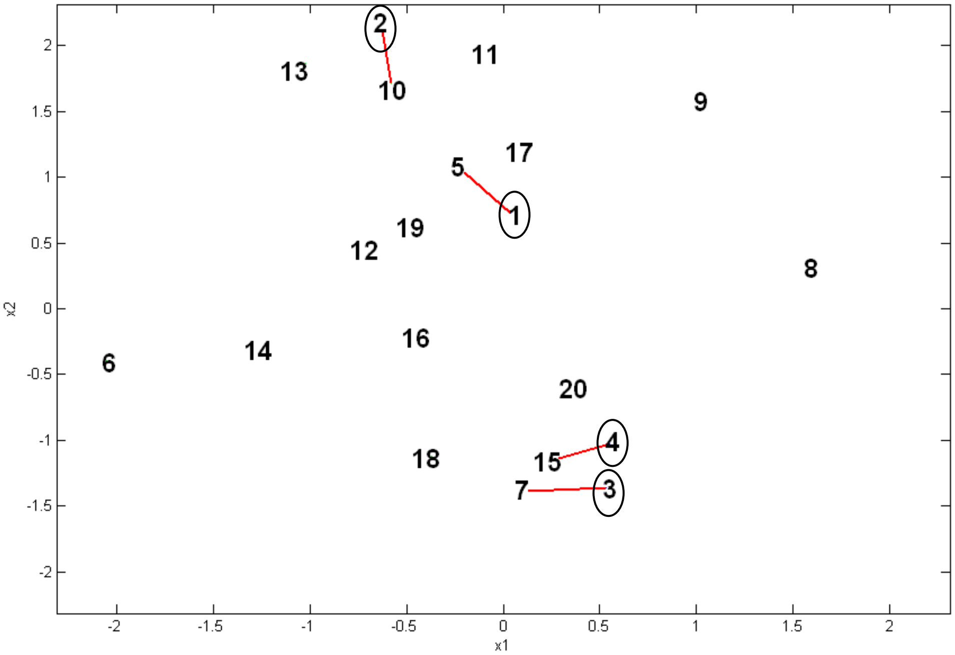

Figure 19. Minimum weight bipartite matching on 20 points; $m = 4$ and $n = 16$ with no change point. The control set is circled; line segments connect observations that are paired in the optimal bipartite matching.

k	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
q_k	1	2	2	3	3	3	3	4	4	4	4	4	4	4	4
Z_k	1	1	2	2	2	3	3	3	3	3	4	4	4	4	4

Table 4. Critical envelope $\mathbf{q}_\alpha$ and BAP test statistic $\mathbf{Z}$ for Figure 19 data. Z_k never exceeds q_k, so the null hypothesis of no change is not rejected.

In contrast, Figure 20 shows the same control data, but the test data are different in that a change point exists at $\tau = 13$, where the index τ is in reference to the pooled data set $\{X_1, \ldots, X_{20}\}$. Specifically, observations $\{X_5, \ldots, X_{12}\}$ are the same as in the no-change case, but $\{X_{13}, \ldots, X_{20}\}$ are translated by 2 units in both dimensions from the no-change case. The control data are circled as before, and the data affected by the change point are circumscribed by a box to clearly show the mean shift.

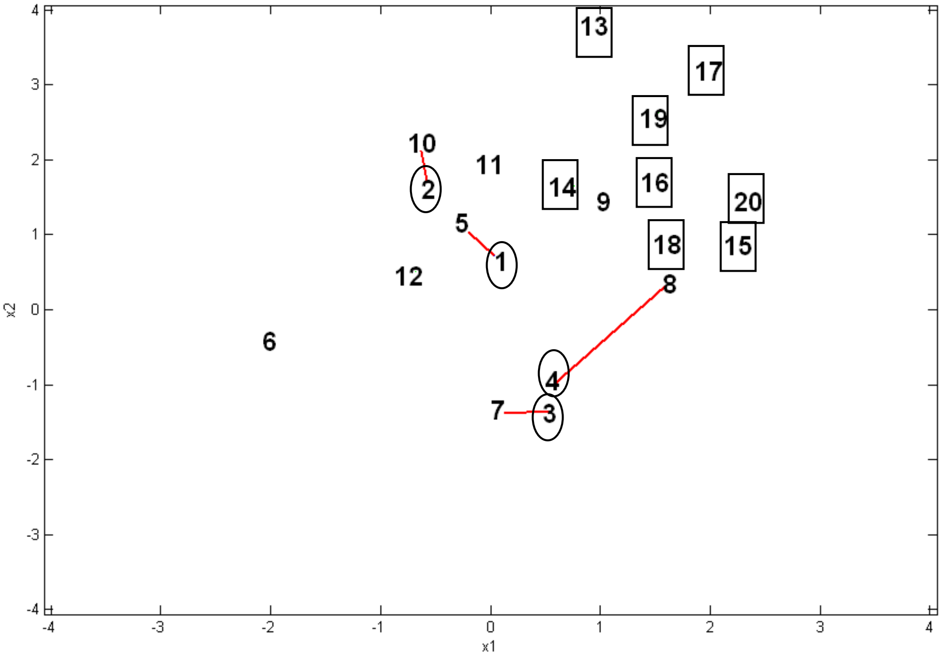

Figure 20. Minimum weight bipartite matching on 20 points; $m = 4$ and $n = 16$ with a change point $\tau = 13$. The control set is circled; post-change point observations are boxed. Line segments connect observations that are paired in the optimal bipartite matching.

k	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
q_k	1	2	2	3	3	**3**	3	4	4	4	4	4	4	4	4
Z_k	1	1	2	3	3	**4**	4	4	4	4	4	4	4	4	4

Table 5. Critical envelope $\mathbf{q}_\alpha$ and BAP test statistic $\mathbf{Z}$ for Figure 20 data. $Z_6 > q_6$, so the null hypothesis of no change is rejected.

Table 5 shows $\mathbf{q}_\alpha$ and test statistic $\mathbf{Z}$ for Figure 20. Since $Z_6 > q_6$, we reject the null hypothesis in favor of the alternative that a change point exists in the test data.

An alternative to the BAP test, as we have formulated it here, is required for the case $m \geq n$, where there exist at least as many control points as test points, since the BAP test would pair every test point to a control point in such cases. A matching-based possibility appeals to previous results: First, assign to each point in the control set some measure of centralness or depth relative to that set; call this measure the observation "quality." Then compute an optimal bipartite matching and let Q_j denote the quality of the control observation that is paired in the matching with test observation X_{m+j}. Since $m \geq n$, every observation in the test set is paired with some observation in the control set, resulting in the sequence $(Q_1, \ldots, Q_n)$. Now assign ranks to this sequence and perform a change-point test on it based on ranks. The existence of a change-point in the rank sequence corresponds to the existence of a change-point in the test set.

The BAP test invites consideration of an ensemble extension similar to the ensemble extension of the NAP test. That is, compute the BAP statistic for each optimal matching in an orthogonal sequence of optimal matchings, and evaluate the collection $\{\mathbf{Z}_1, \mathbf{Z}_2, \ldots, \mathbf{Z}_{N-1}\}$ for additional change-point information. We do not explore this concept here; rather, the proposed BAP test and associated discussion only begin to examine what appear to be rich opportunities to apply bipartite matching ideas to the change-point problem, and we mention them here to inspire further research. In the next chapter, we will examine the performance of our proposed non-bipartite matching methods only, and leave an in-depth study of bipartite methods for future work.

IV. SIMULATION STUDY

A. METHODOLOGY

The usefulness of the tests presented in this paper lies in their power to identify change points under a wide range of alternative hypotheses. In this chapter, we present the results of computer simulations that compare the power of the Sum of Pair Maxima (SPM), Non-Bipartite Accumulated Pairs (NAP), and Ensemble Sum of Pair Maxima (ESPM) tests for a variety of scenarios. We compare these tests to the James, James, and Siegmund (JJS) (1992) test discussed in Section II.B.2. In every case the sample space is $\mathbb{R}^d$ (where *d* is observation dimension), sample size is $N = 200$, and test significance level is $\alpha = 0.05$. The choice $N = 200$ is based on a desire to investigate test behavior for a moderately large sample size while avoiding excessive computation times for large simulations. Detection power is the performance metric, where **power** is defined as the probability of rejecting the null hypothesis when it is false. Each power estimate in the tables of this chapter is the fraction of times that a particular test indicates that a change point has occurred under the given conditions, based on 1000 simulations. We use the Mahalanobis distance function given by (2.16) (estimating *V* by the sample covariance matrix) as a natural choice to determine interpoint costs in $\mathbb{R}^d$ unless otherwise specified. Section C.4 of this chapter shows performance results for cases where different cost functions are considered. In all other cases, the scenarios vary according to the following factors:

1) *Underlying distribution family.* We consider the following probability distribution families:

a) multivariate normal, denoted F_{MVN},

b) a multivariate normal mixture, denoted F_{mix}, as a heavy-tailed case, and

c) a multivariate Weibull distribution, denoted F_{Weib}, as a skewed case.

2) *Dimension*. We evaluate two dimensions: $d = 5$ and $d = 20$.

3) *Change-point location*. We examine cases where the change point occurs in the middle and toward the end of the observation sequence.

4) *Change parameter*. We consider changes in distribution mean and scale.

5) *Type of change*. At a change point, we consider cases where the parameter undergoing change does so in an abrupt (jump) or gradual (drift) manner.

6) *Magnitude of change*. We examine changes of various magnitudes.

Specifically, F_{MVN} is the cumulative distribution associated with density function

$$(4.1)\qquad f_{\text{MVN}}\left(\mathbf{x};\boldsymbol{\mu},\Sigma\right)=\frac{1}{\left(2\pi\right)^{N/2}\left|\Sigma\right|^{1/2}}\exp\left\{-\frac{1}{2}\left(\mathbf{x}-\boldsymbol{\mu}\right)'\Sigma^{-1}\left(\mathbf{x}-\boldsymbol{\mu}\right)\right\}\quad\forall\mathbf{x}\in\mathbb{R}^{d},$$

where $\boldsymbol{\mu}\in\mathbb{R}^{d}$ and $\Sigma\in\mathbb{R}^{d\times d}$ are the distribution mean and covariance matrix, respectively. The in-control data for the multivariate normal case have $\boldsymbol{\mu}=\mathbf{0}$ and $\Sigma = I_d$, where I_d is the $d\times d$ identity matrix. The post-change-point data have a different mean or scale as specified in the next section.

F_{mix} is constructed as follows: Let $U \sim F_{\text{MVN}}$, let Z be a Bernoulli random variable with success probability p_{mix}, and let σ_{mix} be some scalar larger than 1. Then the random variable

$$(4.2)\qquad X=\left[1+\left(\sigma_{mix}-1\right)Z\right]U\sim F_{\text{mix}},$$

and p_{mix} and σ_{mix} are the proportion and scale of the mixing, respectively. For this study, we set $p_{\text{mix}} = 0.1$ and $\sigma_{\text{mix}} = 4$.

F_{Weib} is defined to be the distribution on $\mathbb{R}^d$ associated with d independent identically distributed univariate Weibull random variables; that is,

$$(4.3)\qquad F_{\text{Weib}}\left(\mathbf{x};\eta,\boldsymbol{\beta}\right)=\begin{cases}\prod_{i=1}^{d}\left[1-\exp\left\{-\left(\dfrac{x_i}{\beta_i}\right)^{\eta}\right\}\right], & \mathbf{x}=\left(x_1,\ldots,x_d\right)\in\mathbb{R}_{+}^{d},\\ 0\quad\text{otherwise},\end{cases}$$

where η and β_i are the univariate Weibull shape and scale parameters, respectively, and $\mathbb{R}_+^d$ denotes the closed upper half-space of $\mathbb{R}^d$. For this study we set $\eta = 1.5$. For the in-control data, $\boldsymbol{\beta} = \mathbf{1}$; $\boldsymbol{\beta}$ varies for the post-change-point data as specified in the next section.

We consider a change point at $\tau = 101$ (for a 50-50 percent split of pre- and post-change data) and $\tau = 161$ (for an 80-20 percent split of pre- and post-change data). We examine change magnitudes, denoted by Δ, ranging from 0 to 1.0 by increments of 0.25. For the case $\Delta = 0$ the null hypothesis is *true* and the tabulated power at $\Delta = 0$ is an estimate of the test's **false alarm rate**, which is the likelihood that a test rejects the null hypothesis when it is in fact true. False alarm rate ideally is 0.05 at significance level $\alpha = 0.05$. For the case $\Delta > 0$, Δ indicates the *total magnitude* of the change from the first to last observation. So, for a jump change Δ is magnitude of the abrupt change that occurs at change point τ. For a drift change, the parameter undergoing change varies linearly from its reference level so that the total change magnitude between the first and last observation is Δ. For changes in distribution mean, we simulate the change in the first component of each observation only. For changes in distribution scale, we simulate the change in all components. Figures 21-28 shows cases of mean and scale changes of jump and drift variety for change points at $\tau = 101$ or $\tau = 161$ associated with the multivariate normal case for illustration purposes. The *x*-axis is sequence label *i*; the *y*-axis is the value of the first component of the i^{th} observation.

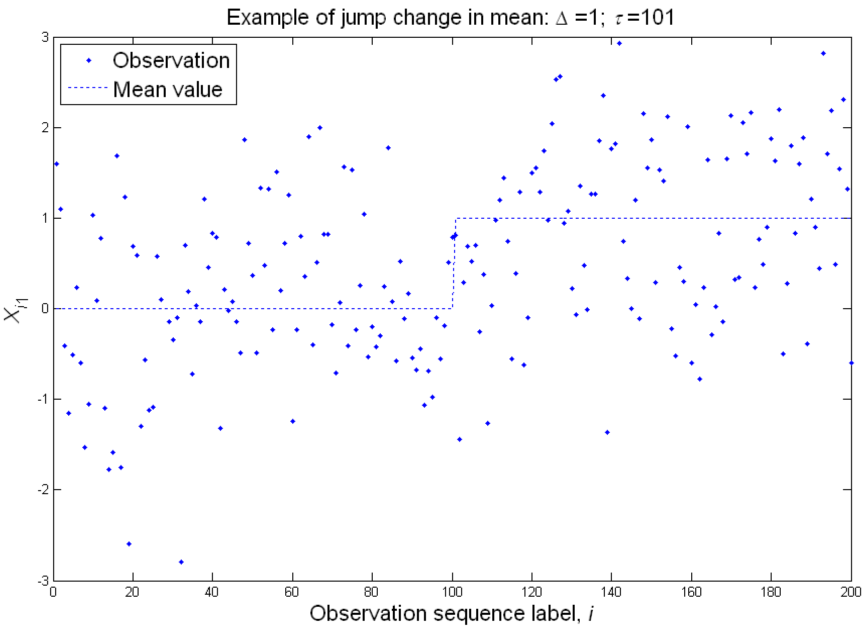

Figure 21. Typical change scenario: mean jump at $\tau = 101$. The mean jumps in the first dimension only and remains fixed in all other dimensions.

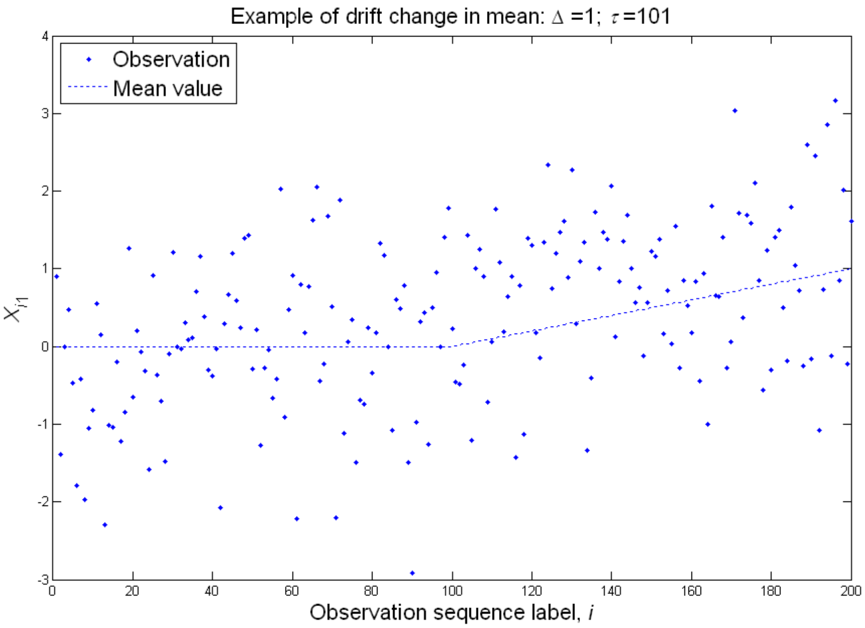

Figure 22. Typical change scenario: mean drift at $\tau = 101$. The mean drifts in the first dimension only and remains fixed in all other dimensions.

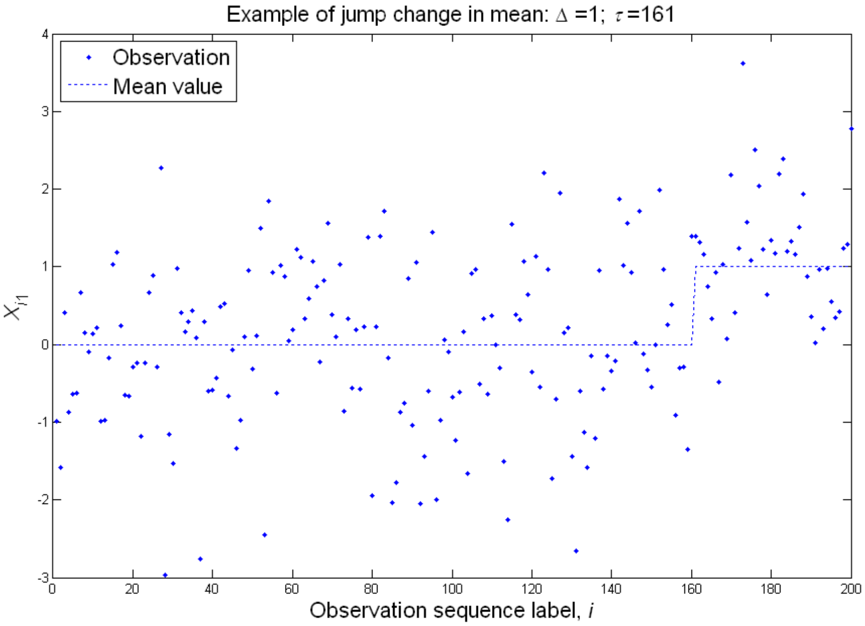

Figure 23. **Typical change scenario: mean jump at $\tau = 161$. The mean jumps in the first dimension only and remains fixed in all other dimensions.**

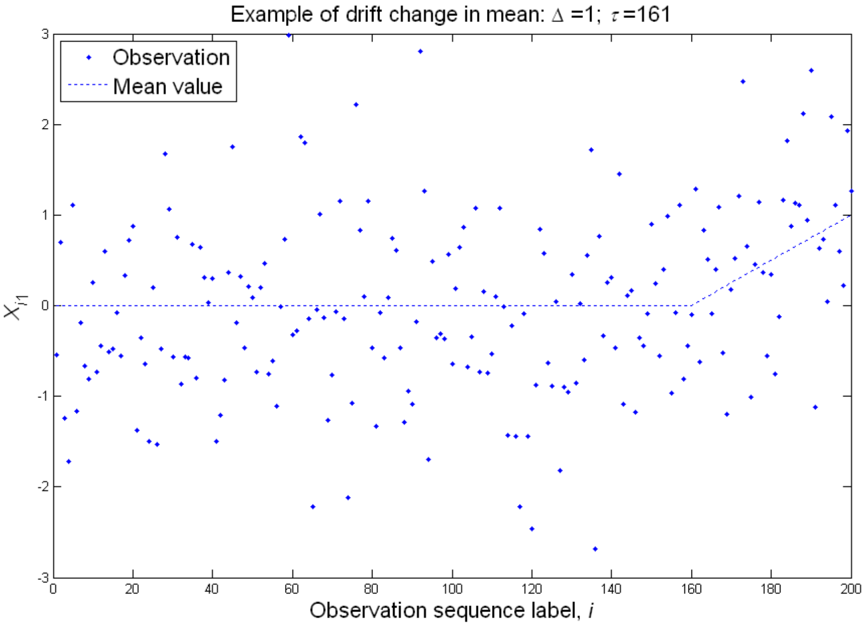

Figure 24. **Typical change scenario: mean drift at $\tau = 161$. The mean drifts in the first dimension only and remains fixed in all other dimensions.**

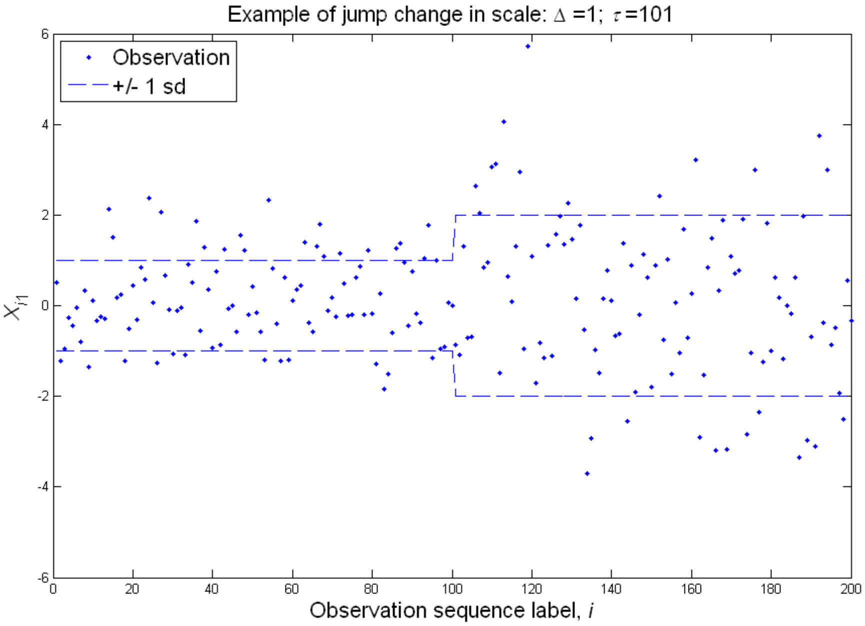

Figure 25. **Typical change scenario: scale jump at $\tau = 101$. The scale jumps in all dimensions.**

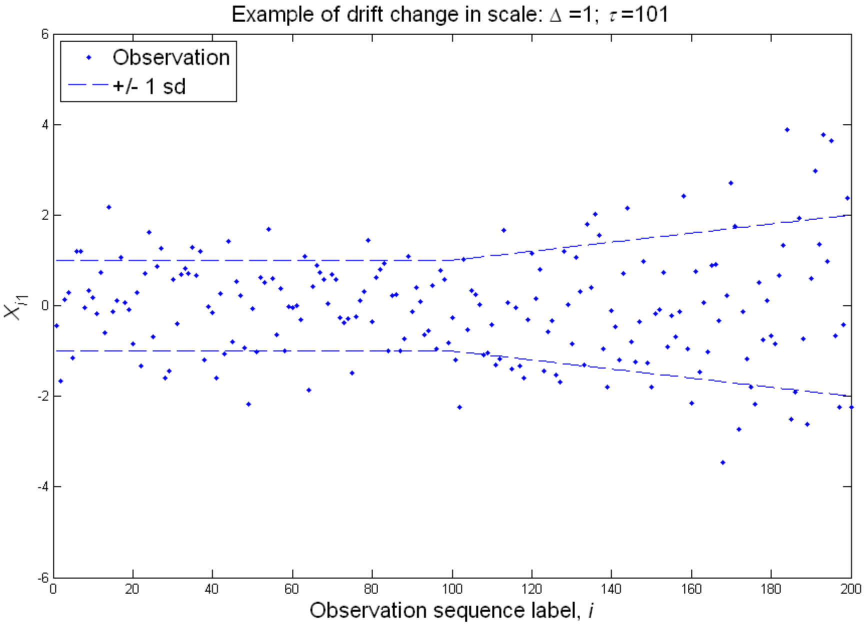

Figure 26. **Typical change scenario: scale drift at $\tau = 101$. The scale jumps in all dimensions.**

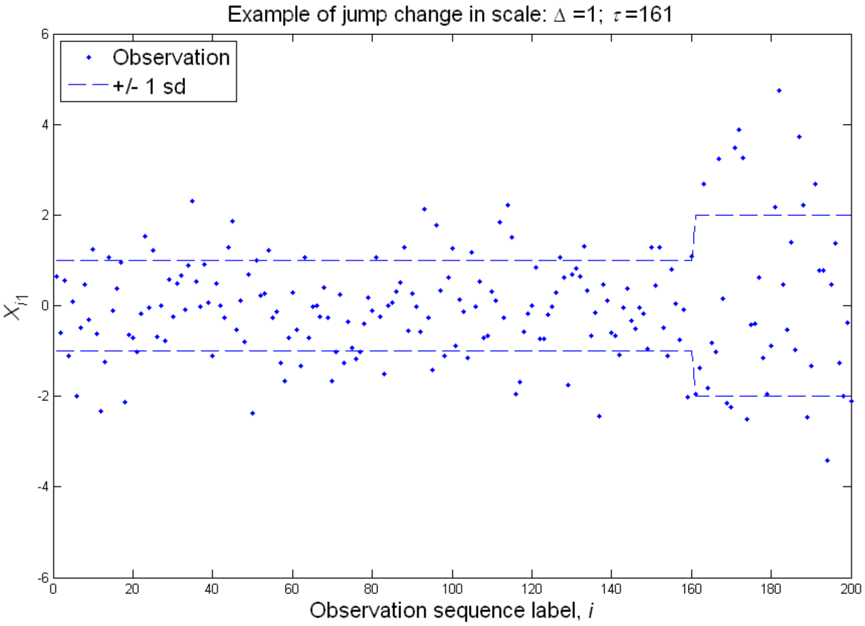

Figure 27. **Typical change scenario: scale jump at $\tau = 161$. The scale jumps in all dimensions.**

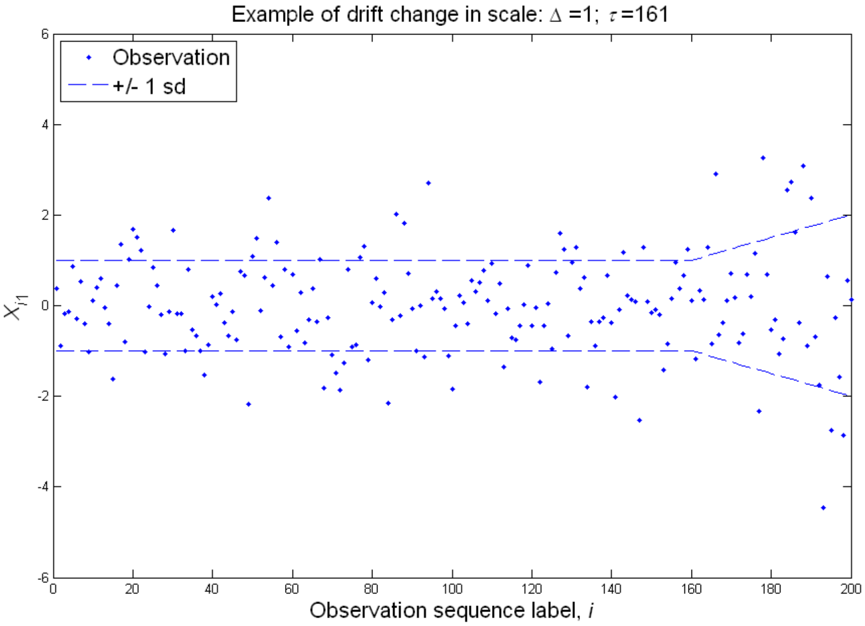

Figure 28. **Typical change scenario: scale drift at $\tau = 161$. The scale jumps in all dimensions.**

B. PERFORMANCE RESULTS

Tables 6-13 show power estimates for the Sum of Pair-Maxima (SPM), Non-Bipartite Accumulated Pairs (NAP), Ensemble Sum of Pair-Maxima (ESPM) and James, James, and Siegmund (JJS) tests for a variety of scenarios at test level $\alpha = 0.05$. Attained significance levels for the SPM and NAP tests with $N = 200$ are $\alpha_{\text{SPM}} = 0.04962$ and $\alpha_{\text{NAP}} = 0.04957$. ESPM critical values are obtained via simulation (see Appendix B). JJS critical values are analytically based on the assumption that the underlying distribution is multivariate normal with a common but unknown covariance matrix. While the JJS test is analytically designed to detect abrupt mean changes only, we observe in our simulations that it is also sensitive to gradual mean changes and scale changes while maintaining a false alarm rate consistent with test significance level. Therefore, we consider the JJS test for comparison in those cases as well.

Each table specifies the distribution, dimensionality, change parameter (mean or scale), and change type (jump or drift) along the top. The left-most column shows the total magnitude of the change for the varying parameter in that case. The change magnitude at *i* is $\Delta_i = \Delta$ for jump changes and $\Delta_i = (i - \tau + 1)\Delta / (N - \tau + 1)$ for drift changes. For a mean change at change point τ, observation *i* is distributed as follows:

$$(4.4)\quad \begin{array}{l} \left.\begin{array}{l} X_i \sim F_{\text{MVN}}\,, \quad i < \tau \\ X_i - (\Delta_i, 0, 0, \ldots, 0) \sim F_{\text{MVN}}\,, \quad i \geq \tau \end{array}\right\} \text{ for the MVN case;} \\[2ex] \left.\begin{array}{l} X_i \sim F_{\text{mix}}\,, \quad i < \tau \\ X_i - (\Delta_i, 0, 0, \ldots, 0) \sim F_{\text{mix}}\,, \quad i \geq \tau \end{array}\right\} \text{ for the mixture case.} \end{array}$$

We model changes in the mean vector by changing only its first component because of the rotational invariance of F_{MVN} and F_{mix}. For a scale change at change point τ, observation *i* is distributed as follows:

$$(4.5)\qquad \left.\begin{array}{l} X_i \sim F_{\text{MVN}}\ , \quad i<\tau \\ \dfrac{X_i}{1+\Delta_i} \sim F_{\text{MVN}}\ , \quad i\ge\tau \end{array}\right\} \text{ for the MVN case;}$$

$$\left.\begin{array}{l} X_i \sim F_{\text{mix}}\ , \quad i<\tau \\ \dfrac{X_i}{1+\Delta_i} \sim F_{\text{mix}}\ , \quad i\ge\tau \end{array}\right\} \text{ for the mixture case.}$$

For fixed η, a change in $\boldsymbol{\beta}$ results in a mean and scale change for the Weibull distribution. We change only the first component of $\boldsymbol{\beta}$ to be consistent with the other two distribution family cases, so observation *i* is distributed as follows:

$$(4.6)\qquad \left.\begin{array}{l} X_i \sim F_{\text{Weib}} \text{ with } \boldsymbol{\beta}=\mathbf{1}, \quad i<\tau \\ X_i \sim F_{\text{Weib}} \text{ with } \boldsymbol{\beta}=\left(1+\Delta_i,1,\ldots,1\right), \quad i\ge\tau \end{array}\right\} \text{ for the MV Weibull case.}$$

1. Multivariate Normal Case

a. Changes in Location

Tables 6-10 present power estimates for the multivariate normal scenarios under different alternatives. Tables 6-8 are associated with a mean change. One would expect the JJS test to be superior to nonparametric tests for the mean jump cases, as JJS is a parametric test based on the assumptions of multivariate normality and a single abrupt change in distribution mean. Our simulation results suggest that the JJS test is superior overall in both the jump and drift cases. However, both the SPM and NAP tests show appreciable power in each case, and even more noteworthy that the power of the ESPM test is comparable to JJS in some cases. In particular, when the change point occurs in the middle of the observation sequence (Tables 6 and 7), the JJS test and ESPM test perform comparably. When the change point is away from the middle of the observation sequence (Table 8) all the tests suffer somewhat, since fewer post-change data are present to indicate a change. However, the nonparametric tests seem to suffer more power loss than the JJS test. Furthermore, the power of all four tests is reduced in higher dimension for changes of fixed magnitude (compare Table 6 with $d=5$ to Table 7 with $d=20$), since the magnitude of the change becomes a smaller fraction of the average distance between points as dimensionality increases. This effect appears to impact the ESPM and JJS tests comparably.

F_{MVN}; $d = 5$; mean change	Jump; $\tau = 101$				Drift; $\tau = 101$			
Δ	SPM	NAP	ESPM	JJS	SPM	NAP	ESPM	JJS
0.00	0.06	0.05	0.04	0.05	0.04	0.05	0.06	0.07
0.25	0.07	0.05	0.18	0.14	0.06	0.05	0.10	0.09
0.50	0.10	0.09	0.60	0.52	0.07	0.05	0.27	0.22
0.75	0.23	0.18	0.93	0.93	0.11	0.09	0.55	0.53
1.00	0.41	0.33	1.00	1.00	0.20	0.16	0.84	0.85

SPM: Sum of Pair-Maxima NAP: Non-Bipartite Accumulated Pairs
ESPM: Ensemble Sum of Pair-Maxima JJS: James, James, and Siegmund test

Table 6. Test power to detect a mean change of magnitude Δ for MVN case with dimension $d = 5$ and change point $\tau = 101$ based on $N = 200$, $\alpha = 0.05$, and 1000 simulations. Jump case is to the left; drift case is to the right.

F_{MVN}; $d = 20$; mean change	Jump; $\tau = 101$				Drift; $\tau = 101$			
Δ	SPM	NAP	ESPM	JJS	SPM	NAP	ESPM	JJS
0.00	0.05	0.05	0.05	0.03	0.05	0.05	0.05	0.04
0.25	0.07	0.06	0.11	0.07	0.05	0.05	0.07	0.04
0.50	0.09	0.07	0.33	0.20	0.07	0.05	0.13	0.09
0.75	0.11	0.09	0.71	0.63	0.08	0.07	0.31	0.23
1.00	0.22	0.16	0.95	0.95	0.11	0.09	0.56	0.49

SPM: Sum of Pair-Maxima NAP: Non-Bipartite Accumulated Pairs
ESPM: Ensemble Sum of Pair-Maxima JJS: James, James, and Siegmund test

Table 7. Test power to detect a mean change of magnitude Δ for MVN case with dimension $d = 20$ and change point $\tau = 101$ based on $N = 200$, $\alpha = 0.05$, and 1000 simulations. Jump case is to the left; drift case is to the right.

F_{MVN}; $d = 5$; mean change	Jump; $\tau = 161$				Drift; $\tau = 161$			
Δ	SPM	NAP	ESPM	JJS	SPM	NAP	ESPM	JJS
0.00	0.05	0.06	0.06	0.06	0.05	0.06	0.05	0.05
0.25	0.06	0.05	0.08	0.08	0.05	0.05	0.06	0.07
0.50	0.07	0.08	0.22	0.30	0.06	0.06	0.10	0.12
0.75	0.12	0.10	0.52	0.72	0.06	0.08	0.15	0.27
1.00	0.19	0.20	0.81	0.96	0.05	0.08	0.28	0.51

SPM: Sum of Pair-Maxima NAP: Non-Bipartite Accumulated Pairs
ESPM: Ensemble Sum of Pair-Maxima JJS: James, James, and Siegmund test

Table 8. Test power to detect a mean change of magnitude Δ for MVN case with dimension $d = 5$ and change point $\tau = 161$ based on $N = 200$, $\alpha = 0.05$, and 1000 simulations. Jump case is to the left; drift case is to the right.

b. Changes in Scale

Tables 9 and 10 show power estimates for multivariate normal scale changes. Note that the case in Table 9 ($d = 5$, $\tau = 101$) varies from the case in Table 10 ($d = 20$, $\tau = 161$) in both dimensionality and change point. We observe that the SPM and NAP tests demonstrate reasonable power to detect scale changes, while the ESPM test shows impressive power to do so. Recall that the JJS test is not specifically designed to detect scale changes; however, it does. Interestingly, JJS exhibits its worst power among these scenarios in lower dimension with a 50-50 split and mean jump, and its best power in higher dimension with an 80-20 split and mean drift.

F_{MVN}; $d = 5$; scale change	Jump; $\tau = 101$				Drift; $\tau = 101$			
Δ	SPM	NAP	ESPM	JJS	SPM	NAP	ESPM	JJS
0.00	0.06	0.05	0.05	0.04	0.05	0.05	0.05	0.05
0.25	0.16	0.11	0.32	0.09	0.08	0.08	0.15	0.14
0.50	0.51	0.42	0.97	0.15	0.27	0.20	0.52	0.27
0.75	0.88	0.88	1.00	0.19	0.52	0.46	0.92	0.42
1.00	0.99	0.99	1.00	0.24	0.79	0.77	1.00	0.54

SPM: Sum of Pair-Maxima NAP: Non-Bipartite Accumulated Pairs
ESPM: Ensemble Sum of Pair-Maxima JJS: James, James, and Siegmund test

Table 9. Test power to detect a scale change of magnitude Δ for MVN case with dimension $d = 5$ and change point $\tau = 101$ based on $N = 200$, $\alpha = 0.05$, and 1000 simulations. Jump case is to the left; drift case is to the right.

F_{MVN}; $d = 20$; scale change	Jump; $\tau = 161$				Drift; $\tau = 161$			
Δ	SPM	NAP	ESPM	JJS	SPM	NAP	ESPM	JJS
0.00	0.05	0.06	0.05	0.04	0.05	0.05	0.06	0.04
0.25	0.08	0.08	0.24	0.26	0.06	0.07	0.10	0.23
0.50	0.25	0.28	0.83	0.66	0.09	0.11	0.25	0.70
0.75	0.56	0.75	1.00	0.90	0.17	0.25	0.55	0.93
1.00	0.85	0.99	1.00	0.98	0.28	0.49	0.86	0.99

SPM: Sum of Pair-Maxima NAP: Non-Bipartite Accumulated Pairs
ESPM: Ensemble Sum of Pair-Maxima JJS: James, James, and Siegmund test

Table 10. Test power to detect a scale change of magnitude Δ for MVN case with dimension $d = 5$ and change point $\tau = 161$ based on $N = 200$, $\alpha = 0.05$, and 1000 simulations. Jump case is to the left; drift case is to the right.

2. Multivariate Normal Mixture Case

a. Changes in Location

Table 11 shows power results for the case of an underlying multivariate normal mixture with $p_{\text{mix}} = 0.10$ and $\sigma_{\text{mix}} = 4$ as an example of a heavy-tailed case:

$F_{\text{mix}}; d = 5;$ mean change	Jump; $\tau = 101$				Drift; $\tau = 101$			
Δ	SPM	NAP	ESPM	JJS	SPM	NAP	ESPM	JJS
0.00	0.05	0.05	0.04	0.27	0.04	0.04	0.06	0.28
0.25	0.07	0.05	0.14	0.28	0.07	0.06	0.09	0.26
0.50	0.09	0.08	0.56	0.38	0.07	0.07	0.21	0.33
0.75	0.20	0.15	0.88	0.61	0.10	0.09	0.47	0.39
1.00	0.36	0.25	0.99	0.85	0.15	0.12	0.76	0.55

SPM: Sum of Pair-Maxima NAP: Non-Bipartite Accumulated Pairs
ESPM: Ensemble Sum of Pair-Maxima JJS: James, James, and Siegmund test

Table 11. Test power to detect a mean change of magnitude Δ for MVN mixture case with dimension $d = 5$ and change point $\tau = 101$ based on $N = 200$, $\alpha = 0.05$, and 1000 simulations. Jump case is to the left; drift case is to the right. Shading indicates excessive false alarm rate.

The matching-based tests demonstrate results comparable to their respective powers in the similar multivariate normal case (compare to Table 6) and they retain a false alarm rate consistent with test significance level. However, the false alarm rate for the JJS test far exceeds 5% and therefore disqualifies it for comparison at the 0.05 test level. We explore this phenomenon in a separate study using 10,000 simulations ($N = 200$, 50-50 split) and find that the JJS false alarm rates exceed test level even for small mixing probabilities. The problem gets worse as dimensionality increases, as shown in Figure 29.

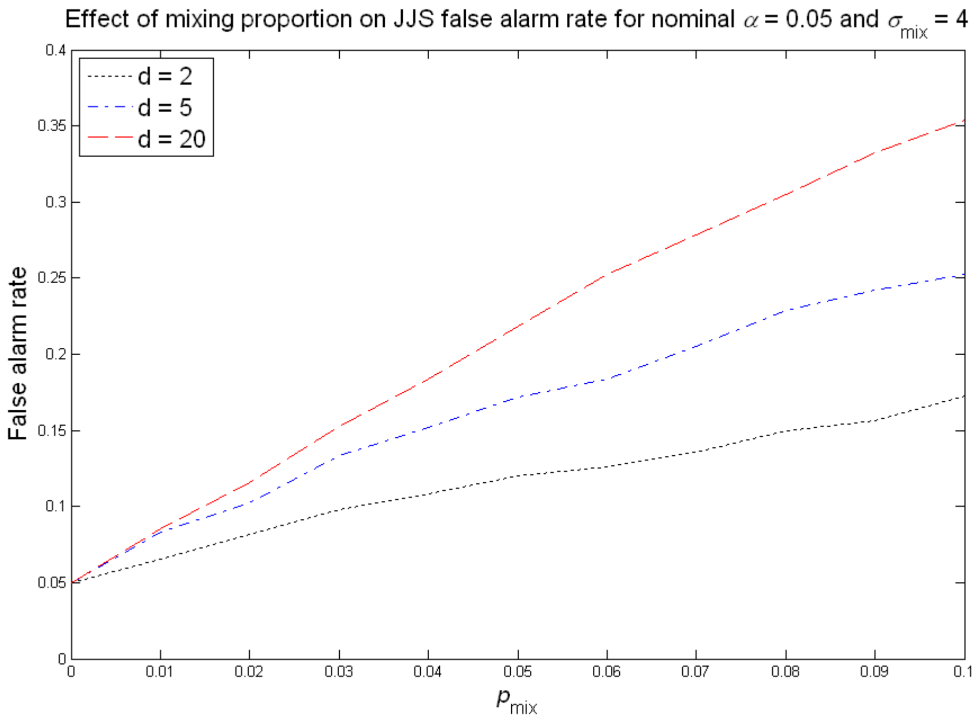

Figure 29. Effect of mixing proportion on JJS false alarm rate for MVN mixture cases with dimension $d = 2$, 5, and 20 based on $N = 200$, $\alpha = 0.05$, and 10,000 simulations. Scale of mixing $\sigma_{mix} = 4$; proportion of mixing p_{mix} varies along the *x*-axis.

b. Changes in Scale

As in the multivariate normal case, Table 12 shows that the SPM and NAP tests have fair power to detect scale changes and the ESPM test has noteworthy power to do so. Again, excessive false alarm rate makes JJS an unacceptable test for these scenarios. These multivariate normal mixture cases of changing location and scale highlight the utility of nonparametric change-point approaches such as the SPM, NAP, and ESPM tests in that they are not limited by strict distributional assumptions.

F_{mix}; $d = 5$; scale change	Jump; $\tau = 101$				Drift; $\tau = 101$			
Δ	SPM	NAP	ESPM	JJS	SPM	NAP	ESPM	JJS
0.00	0.05	0.05	0.05	0.25	0.06	0.06	0.05	0.29
0.25	0.12	0.09	0.31	0.29	0.08	0.08	0.13	0.32
0.50	0.38	0.28	0.92	0.31	0.16	0.13	0.48	0.39
0.75	0.75	0.69	1.00	0.32	0.35	0.27	0.85	0.45
1.00	0.93	0.92	1.00	0.32	0.54	0.45	0.99	0.50

SPM: Sum of Pair-Maxima NAP: Non-Bipartite Accumulated Pairs
ESPM: Ensemble Sum of Pair-Maxima JJS: James, James, and Siegmund test

Table 12. Test power to detect a scale change of magnitude Δ for MVN mixture case with dimension $d = 5$ and change point $\tau = 101$ based on $N = 200$, $\alpha = 0.05$, and 1000 simulations. Jump case is to the left; drift case is to the right. Shading indicates excessive false alarm rate.

3. Multivariate Weibull Case

Table 13 presents power results associated with the multivariate Weibull distribution as an example of a skewed case. For these simulations shape parameter $\eta = 1.5$ remains fixed while scale parameter $\boldsymbol{\beta}$ varies from 1 to 2 in the first dimension only; this corresponds to coincident changes in both location and scale. While false alarm rates for the JJS test are not as excessive for this case as for the multivariate mixture case, they still clearly violate the specified 0.05 test level. As before, the matching-based tests appear to respect test level.

F_{Weib}; $d = 5$; β change	Jump; $\tau = 101$				Drift; $\tau = 101$			
Δ	SPM	NAP	ESPM	JJS	SPM	NAP	ESPM	JJS
0.00	0.05	0.05	0.06	0.09	0.06	0.05	0.05	0.09
0.25	0.08	0.06	0.25	0.22	0.06	0.07	0.12	0.17
0.50	0.13	0.10	0.70	0.70	0.08	0.07	0.35	0.46
0.75	0.25	0.20	0.95	0.96	0.15	0.12	0.70	0.81
1.00	0.34	0.27	0.99	1.00	0.23	0.20	0.86	0.94

SPM: Sum of Pair-Maxima NAP: Non-Bipartite Accumulated Pairs
ESPM: Ensemble Sum of Pair-Maxima JJS: James, James, and Siegmund test

Table 13. Test power to detect a change in the scale parameter β of magnitude Δ for MV Weibull case with dimension $d = 5$ and change point $\tau = 101$ based on $N = 200$, $\alpha = 0.05$, and 1000 simulations. Jump case is to the left; drift case is to the right. Shading indicates excessive false alarm rate.

In summary, the Sum of Pair-Maxima (SPM), Non-Bipartite Accumulated Pairs (NAP), and Ensemble Sum of Pair-Maxima (ESPM) tests all demonstrate power to detect a change point in every examined case for different underlying different distributions, dimensionality, change-point location, change parameter, and type of change while achieving a significance level consistent with nominal test level. The power of each test is reduced as dimension increases or as change-point location moves away from the middle of the observation sequence.

The ESPM test outperforms the SPM and NAP tests in every case and is preferable among the three tests for use. The ESPM test has power comparable to the parametric JJS test in the case of a mean change when the underlying distribution is multivariate normal, except perhaps when the change-point is far away from the center of the sequence. The ESPM test is preferable to the JJS test in non-normal cases due to excessive JJS false alarm rates, and is superior in detecting scale changes when the underlying distribution is normal. The NAP test should be considered for use if one desires information about the location of a change point in addition to detecting whether or not a change-point exists.

C. DIFFERENT COST FUNCTIONS

To gain insight regarding the impact of cost function selection we compare the performance of the ESPM test using Mahalanobis distance (MD), Euclidean distance (ED), Mahalanobis distance, robust (MD-R), and multivariate rank distance (RD) as defined in (2.15)-(2.23) for a few representative cases. We list MD in the first column as the reference cost measure which was used for all previous cases in the simulation study. For MD-R we set the nearest-neighbor parameter *k* (as identified in the discussion preceding equation (2.17)) equal to 8, which is in the range of recommended values for that parameter given by Wang and Raferty (2002).

In every case, MD and MD-R performance are nearly identical, and ED and RD performance are nearly identical. ED and RD perform as well or better than MD and MD-R; this performance difference is more evident in higher dimension and is most evident in for the multivariate Weibull case. These differences seem attributable to the fact that MD and MD-R must estimate the covariance of the underlying distribution, while for the cases we examine the underlying covariance is very close to the identity covariance assumed by ED and RD.

F_{MVN}; ESPM; mean change	$d = 5$; jump; $\tau = 101$				$d = 20$; jump; $\tau = 101$			
Δ	MD	ED	MD-R	RD	MD	ED	MD-R	RD
0.00	0.06	0.06	0.05	0.06	0.05	0.05	0.05	0.05
0.25	0.16	0.16	0.16	0.16	0.11	0.12	0.11	0.12
0.50	0.56	0.58	0.56	0.57	0.31	0.36	0.30	0.36
0.75	0.93	0.94	0.93	0.95	0.72	0.80	0.71	0.79
1.00	1.00	1.00	1.00	1.00	0.96	0.99	0.96	0.99

MD: Mahalanobis distance
ED: Euclidean distance
MD-R: Mahalanobis distance, robust
RD: Multivariate rank distance

Table 14. Ensemble Sum of Pair-Maxima (ESPM) test power to detect a mean change of magnitude Δ for MVN case with dimension $d = 5$ and change point $\tau = 101$ under different cost functions, based on $N = 200$, $\alpha = 0.05$, and 1000 simulations. Jump case is to the left; drift case is to the right.

F_{mix}; ESPM; mean change	$d = 5$; jump; $\tau = 101$				$d = 20$; jump; $\tau = 101$			
Δ	MD	ED	MD-R	RD	MD	ED	MD-R	RD
0.00	0.05	0.05	0.05	0.05	0.06	0.05	0.06	0.06
0.25	0.16	0.15	0.15	0.16	0.10	0.09	0.10	0.09
0.50	0.46	0.50	0.49	0.52	0.26	0.34	0.28	0.33
0.75	0.89	0.91	0.90	0.90	0.57	0.69	0.62	0.69
1.00	0.99	0.99	0.99	0.99	0.86	0.95	0.91	0.95

MD: Mahalanobis distance
ED: Euclidean distance
MD-R: Mahalanobis distance, robust
RD: Multivariate rank distance

Table 15. Ensemble Sum of Pair-Maxima (ESPM) test power to detect a mean change of magnitude Δ for MVN mixture case with dimension $d = 5$ and change point $\tau = 101$ under different cost functions, based on $N = 200$, $\alpha = 0.05$, and 1000 simulations. Jump case is to the left; drift case is to the right.

F_{Weib}; ESPM; β change	$d = 5$; jump; $\tau = 101$				$d = 20$; jump; $\tau = 101$			
Δ	MD	ED	MD-R	RD	MD	ED	MD-R	RD
0.00	0.05	0.06	0.05	0.05	0.05	0.04	0.05	0.04
0.25	0.24	0.28	0.25	0.28	0.14	0.19	0.15	0.18
0.50	0.70	0.77	0.71	0.76	0.46	0.66	0.47	0.67
0.75	0.93	0.97	0.95	0.97	0.73	0.95	0.77	0.95
1.00	0.99	1.00	1.00	1.00	0.93	1.00	0.95	1.00

MD: Mahalanobis distance
ED: Euclidean distance
MD-R: Mahalanobis distance, robust
RD: Multivariate rank distance

Table 16. Ensemble Sum of Pair-Maxima (ESPM) test power to detect a change in the scale parameter β of magnitude Δ for MV Weibull case with dimension $d = 5$ and change point $\tau = 101$ under different cost functions, based on $N = 200$, $\alpha = 0.05$, and 1000 simulations. Jump case is to the left; drift case is to the right.

D. COMPUTATIONAL DETAILS

Simulations in Section B of this chapter were performed using R version 2.9.0 on the Hamming cluster of the Naval Postgraduate School's High Performance Computing Center (HPC), which is a Sun Microsystems 6048 Blade modular system with 1152 processing cores. We computed non-bipartite weighted matchings using Kolmogorov's (2009) Blossom V algorithm. In its published form, Kolmogorov's algorithm computes a single optimal non-bipartite matching on a set of N observations. We modified this routine slightly in the source language so that it computes a full sequence of orthogonal successive optimal matchings that rather than just a single matching. Table 17 shows typical realized runtimes to compute $N/2$ orthogonal successive optimal matchings calling Kolmogorov's routine in R for various sample sizes.

N	Run time (sec)
20	<0.01
50	0.01
100	0.05
200	0.60
300	2.9
400	11
500	48

Table 17. Typical time to compute $N/2$ orthogonal successive optimal matchings calling Kolmogorov's algorithm (modified to compute orthogonal successive optimal matchings) in R.

We realized significant reductions (at least two orders of magnitude) in total simulation time by taking advantage of the HPC's batch job scheduling capability.

Simulations in Section C were performed using R version 2.9.0 on a Windows XP machine with an Intel ® Pentium ® 4 3.4 GHz processor. We computed non-bipartite weighted matchings using Derigs's (1988) algorithm. The algorithm itself is in the FORTRAN programming language; compiled code is embedded in a dynamic link library file that can be called directly in R, provided by Professor Bo Lu at the Ohio State University. In its current form this algorithm requires that edge weights be integer-valued. For our purposes, we accommodated this requirement by rounding non-integer costs to four decimal places and then scaling each cost by 10^4. Additionally, this routine requires the assignment of a prohibitive cost to the diagonal of any cost matrix to block the pairing of an observation with itself. For this study we set this prohibitive cost to be $N/2$ times the maximum of all interpoint costs. We used this same prohibitive cost for blocking to compute orthogonal successive optimal matchings. Table 18 shows typical realized runtimes to compute $N/2$ orthogonal successive optimal matchings using Derigs's algorithm in R for various sample sizes.

N	Run time (sec)
20	<0.01
50	0.02
100	0.2
200	2.9
300	16
400	56
500	143

Table 18. Typical time to compute $N/2$ orthogonal successive optimal matchings using Derigs's algorithm in R.

As mentioned previously, the theoretical runtime for existing algorithms to find a single optimal non-bipartite matching on a complete graph is $O\left(N^3\right)$. Our ESPM statistic involves computing $N/2$ successive optimal matchings on a graph, which can lead to lengthy run times for large sample sizes. Long runtimes for cases of very large sample size pose a practical limitation to the matching methods we propose. We discuss related research opportunities in the next chapter.

V. CONCLUSIONS AND OPPORTUNITIES FOR FURTHER RESEARCH

In this dissertation, we introduce new nonparametric matching-based approaches to the multidimensional change-point problem. These approaches lead to effective change-point detection procedures and highlight the potential value of matching techniques to more general statistical applications. Our review of the broad field of change-point detection reveals that this continues to be an area of active research and that robust multivariate approaches to this problem remain few. Most existing approaches make restrictive distributional assumptions (such as multivariate normality) or are limited to the single-test case where the potential change point is pre-determined and the problem is the classical one of testing whether two samples of observations are from the same distribution.

We propose four new change-point tests: the Sum of Pair-Maxima (SPM) test, the Non-Bipartite Accumulated Pairs (NAP) test, the Ensemble Sum of Pair-Maxima (ESPM) test, and the Bipartite Accumulated Pairs (BAP) test. The first three tests, designed to test for homogeneity among multivariate data when no observation history is available, all demonstrate power to detect a change point under a variety of alternative hypotheses at fixed false alarm rates. The ESPM test utilizes additional change-point information available from many good (that is, low-weight) orthogonal matchings, and is superior among these nonparametric tests to detect a change point. Additionally, the ESPM test has power comparable to a parametric competitor, the JJS test, even when its parametric assumptions are met. The power of the ESPM test not only establishes itself as an effective change-point test, but also validates matching as a useful approach to the change-point problem.

This research invites several possibilities for extension. One obvious question is whether or not any of these tests might be reasonably extended as sequential change-point tests. While it is difficult in general to sequentialize hypothesis tests, sequential change-point detection techniques would have valuable application. One requirement for such an extension would be to extend the theory presented here as necessary for sequentialization. Another practical problem associated with such an extension is the question of how to efficiently update an existing optimal matching on N observations with the addition of one or more data points to the observation set.

Other opportunities include finding ensemble extensions for the NAP and BAP tests. The fact that the exact distributions of these individual tests are known might make the task of finding exact associated ensemble distributions (or good approximations) more tractable. Additionally, the performance of the BAP test remains to be evaluated.

One challenge to research in this area is the scarcity of non-bipartite weighted matching software modules for typical statistical software applications. The simulation study for this research relies heavily on interfaces between C, C++, or FORTRAN; and S-PLUS®, R, or MATLAB® that we or others have built manually. The mainstreaming of any such interface would greatly enable broader related research. Research opportunities exist to improve the efficiency of non-bipartite matching algorithms. Even using existing algorithms, time improvements would be gained by reducing the number of orthogonal successive optimal matchings computed for the ESPM test. As presented here, the ESPM statistic is formed by summing over $N/2$ orthogonal successive optimal matchings where N is the sample size. Additional research is necessary to determine whether fewer (perhaps far fewer) orthogonal successive optimal matchings are adequate to achieve good detection power against alternate hypotheses. Also, it would be worthwhile to investigate the usefulness of "greedy" algorithms in this context. A greedy matching algorithm finds a good matching on N observations by pairing the two closest points, then the next two closest, and so on until a maximal matching has been constructed. This faster algorithm ($O\left(N^2\right)$) does not in general provide an optimal non-bipartite matching. However, a greedy matching may still be good enough to provide valuable change-point information; we believe this would be a worthwhile area for study.

Rosenbaum's (2005) case for the consistency of the cross-match statistic seems quite reasonable, but as he states it is "admittedly informal." His argument is also constrained to less general alternative hypotheses than we have considered in our work. Because our argument for the consistency of the SPM test (and therefore for the ESPM test by direct extension) requires the consistency of the cross-match statistic, work needs to be done to establish the consistency of the cross-match statistic more formally and against alternative hypotheses of a more general nature.

We alluded previously to the fact that machine health diagnosis and prognosis problems were an initial motivation for this research, and we are interested in ways to apply this work to that area. Such problems are often characterized by high dimensionality and serial correlation. In addition to detecting the *presence* of a change point in a sequence of observations, it would be useful also to estimate *where* in the sequence the change point occurred. Furthermore, it would be helpful in the event that a change point is detected to characterize the nature of the change (for example, abrupt or gradual) and the severity of the change for prognostic purposes such as estimating remaining useful life.

An idea that seems worthy of consideration is a generalization of our matching approach to vertex groupings of cardinality greater than two. The tests we propose here are all matching-based, where we mean matching in the strict graph-theoretical sense as defined in Chapter II. Each matching is a collection of single edges, and each edge is in turn is a two-element subset of vertices. Algorithms to find optimal non-bipartite weighted matchings already exist, and we have demonstrated that matchings can be used for effective statistical inference. However, it might be worth examining whether collections of more than one edge (that is, collections of more that two vertices) might provide useful (or even better) information to the change-point problem. For example, instead of computing an optimal non-bipartite matching on a set of observations one might compute an optimal "three-grouping," where the objective function for optimality might be to minimize the collective cycle cost or collective minimum spanning tree cost across subgroups of size three. Similar to the SPM test, one might consider the sum of group-maxima (or -minima, or -median, or some other unary set operator). Even more general "*k*-groupings" might be considered. Unlike the well-known method of *k*-means clustering, which partitions *N* observations into *k* groups (perhaps of different sizes) based on an objective criterion such as minimizing the sum if within-cluster differences, a "*k*-groupings" approach would specify group size *k* first and then collect vertices into groups so as to minimize some particular criterion. We are not aware of any specialized algorithms to find such groupings and the associated statistical properties of such groupings are likely quite complicated, but these ideas might constitute fertile ground for research.

Another interesting idea involves retaining some of the original observation information for the computation of a test statistic. The methods we propose use the observed data in two distinct steps: First we compute an optimal non-bipartite matching based on observation content excluding data sequence labels, then we compute a nonparametric test statistic based only on sequence labels with respect to that matching. However, it might be useful to carry over additional information from the data into the computation of a test statistic. For example, one might associate with each pair in the matching some measure of pair "quality" based on the cost of the pair. These quality values might then be used as weightings in the computation of a sum of pair-maxima-type statistic, and perhaps improve the detection power of such a test.

Finally, an area upon which we have only touched briefly involves the choice of cost function. In research such as ours the existence of some appropriate dissimilarity measure associated with the sample space of interest is often assumed and from there the desired analysis proceeds. While our theoretical results regarding non-ensemble null distributions depend only on the exchangeability of sequence labels and not on choice of cost function, we expect detection power against alternative hypotheses to depend on that choice. While Mahalanobis distance (or some robust modification of Mahalanobis distance) is a natural choice of cost function for continuous random variables, cases of interest may include a mixture of categorical, ordinal, and continuous random variables. Even for continuous cases, the ability to detect change points with respect to a Mahalanobis distance function might be improved. For example, shifting observations by a component-wise smoothed mean can lead to better covariance matrix estimation in cases where a change point exists. In any case, further study of the effects of cost function choice on the power of tests presented here would be of useful, especially for cases that include categorical or ordinal dimensions. In particular, for real-world application of these methods it would be worth investigating which cost functions lead to the most attractive power characteristics for the specific case at hand.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX A: HIGHER MOMENT DERIVATIONS

In this appendix, we derive various moments associated with the Sum of Pair-Maxima (SPM) statistic.

A. MEAN AND VARIANCE OF Y_1

For a non-bipartite match on $N = 2n$ observations, each of the $(2n)!$ possible assignments of ranks is equally likely under the null hypothesis, and the random variables $Y_1, \ldots, Y_n$ are exchangeable. Therefore, Y_1 takes on the value *t* when observation *t* is paired with some observation *s* of lesser sequence label and (s,t) is indexed as the first among the *n* pairs. But

$$P\{t \text{ is paired with } s\} = \frac{1}{2n-1} \tag{E.1}$$

and

$$P\{(s,t) \text{ is indexed first among matches} | t \text{ is paired with } s\} = \frac{1}{n}, \tag{E.2}$$

so

$$\begin{aligned} P(Y_1 = t) &= \sum_{s=1}^{t-1} \begin{bmatrix} P\{(s,t) \text{ is indexed first among matches } | t \text{ is paired with } s\} \\ \times P\{t \text{ is paired with } s\} \end{bmatrix}, \\ &= \sum_{s=1}^{t-1} \frac{1}{n} \cdot \frac{1}{(2n-1)} = \frac{t-1}{n(2n-1)} = (t-1)\binom{2n}{2}^{-1}, t = 2, \ldots, 2n \end{aligned} \tag{E.3}$$

The desired expressions follow directly:

$$\begin{aligned} E[Y_1] &= \sum_{t=2}^{2n} t \frac{(t-1)}{n(2n-1)} = \frac{2(2n+1)}{3}, \\ E[Y_1^2] &= \sum_{t=2}^{2n} t^2 \frac{(t-1)}{n(2n-1)} = \frac{(2n+1)(3n+1)}{3}, \\ Var[Y_1] &= E[Y_1^2] - E[Y_1]^2 = \frac{(2n+1)(n-1)}{9}. \end{aligned} \tag{E.4}$$

B. COVARIANCE OF Y_1 AND Y_2

First we find $P\left(Y_2 = t \mid Y_1 = s\right)$ by observing that for the case $t < s$,

(E.5)
$$\begin{aligned} P\left(Y_2 = t \mid Y_1 = s\right) &= \frac{\sum_{u=1}^{t-1}(t-2)\binom{2n-2}{2}^{-1}\binom{2n}{2}^{-1} + \sum_{u=t+1}^{s-1}(t-1)\binom{2n-2}{2}^{-1}\binom{2n}{2}^{-1}}{(s-1)\binom{2n}{2}^{-1}} \\ &= \frac{(t-1)(s-3)}{(s-1)(n-1)(2n-3)}, \quad t = 2,\ldots,s-1 \end{aligned}$$

and for the case $t > s$,

(E.6)
$$\begin{aligned} P\left(Y_2 = t \mid Y_1 = s\right) &= \frac{\sum_{u=1}^{s-1}(t-3)\binom{2n-2}{2}^{-1}\binom{2n}{2}^{-1}}{(s-1)\binom{2n}{2}^{-1}} \\ &= \frac{t-3}{(n-1)(2n-3)}, \qquad t = s+1,\ldots,2n. \end{aligned}$$

Now condition on Y_1 to compute $E\left[Y_2 Y_1\right]$:

(E.7)
$$\begin{aligned} E\left[Y_2 \mid Y_1 = s\right] &= \sum_{t=2}^{s-1} t\frac{(t-1)(s-3)}{(s-1)(n-1)(2n-3)} + \sum_{t=s+1}^{2n} t\frac{t-3}{(n-1)(2n-3)} \\ &= \left(\frac{s(s-2)(s-3)}{3(n-1)(2n-3)}\right) + \left(\frac{4n(2n+1)(n-2)}{3} - \frac{s(s+1)(s-4)}{3}\right) \\ &= \frac{4n(2n+1)(n-2) - 2s(s-5)}{3(n-1)(2n-3)} \end{aligned}$$

so

(E.8)
$$\begin{aligned} E\left[Y_2 Y_1\right] &= E\left[Y_1 E\left[Y_2 \mid Y_1\right]\right] \\ &= \sum_{s=2}^{2n} s\frac{4n(2n+1)(n-2) - 2s(s-5)}{3(n-1)(2n-3)} \cdot \frac{(s-1)}{n(2n-1)} \\ &= \frac{8(2n+1)(5n+2)}{45}. \end{aligned}$$

Therefore,

$$
\begin{aligned}
\operatorname{Cov}\left[Y_1,Y_2\right] &= E\left[Y_2\,Y_1\right]-E\left[Y_2\right]E\left[Y_1\right] \\
&= \frac{8(2n+1)(5n+2)}{45}-\left(\frac{2(2n+1)}{3}\right)^2 \qquad \text{(E.9)} \\
&= -\frac{4(2n+1)}{45}.
\end{aligned}
$$

C. THIRD MOMENTS OF Y_1, Y_2, AND Y_3

As in the first and second moment calculations, compute $E\left[Y_1^3\right]$ directly:

(E.10) $$E\left[Y_1^3\right]=\sum_{t=2}^{2n}t^3\frac{(t-1)}{n(2n-1)}=\frac{(2n+1)\left(24n^2+15n+1\right)}{15}.$$

Now compute $E\left[Y_1^2Y_2\right]$ by conditioning on Y_1 using (E.7):

$$
\begin{aligned}
E\left[Y_1^2Y_2\right] &= E\left[Y_1^2E\left[Y_2 \mid Y_1\right]\right] \\
&= \sum_{s=2}^{2n}s^2\frac{4n(2n+1)(n-2)-2s(s-5)}{3(n-1)(2n-3)}\cdot\frac{(s-1)}{n(2n-1)} \qquad \text{(E.11)} \\
&= \frac{4(2n+1)\left(15n^2+10n+1\right)}{45}.
\end{aligned}
$$

In a similar fashion, we apply a series of conditioning arguments to compute $E\left[Y_1\,Y_2\,Y_3\right]$. First, let $Z_{(1)},Z_{(2)},\text{ and }Z_{(3)}$ take on the values of $Y_1,Y_2,\text{ and }Y_3$ such that $Z_{(1)}<Z_{(2)}<Z_{(3)}$ (these $Z_{(i)}$ are unrelated to the Z_i of the BAP test). Then $E\left[Y_1\,Y_2\,Y_3\right]=E\left[Z_{(1)}Z_{(2)}Z_{(3)}\right]=E\left[Z_{(3)}E\left[Z_{(2)}E\left[Z_{(1)}\middle|Z_{(2)},Z_{(3)}\right]\middle|Z_{(3)}\right]\right]$. A direct combinatorial argument gives

(E.12) $$P\left(Z_{(1)}=t_1,Z_{(2)}=t_2,Z_{(3)}=t_3\right)=\frac{6(t_1-1)(t_2-3)(t_3-5)}{\binom{2n}{2}\binom{2n-2}{2}\binom{2n-4}{2}},\quad 2\le t_1<t_2<t_3\le 2n,$$

so

$$
\begin{aligned}
P\left(Z_{(2)}=t_2, Z_{(3)}=t_3\right) &= \sum_{t_1=2}^{2n-2} P\left(Z_{(1)}=t_1, Z_{(2)}=t_2, Z_{(3)}=t_3\right) \\
&= \frac{3\left(t_2-1\right)\left(t_2-2\right)\left(t_2-3\right)\left(t_3-5\right)}{\binom{2n}{2}\binom{2n-2}{2}\binom{2n-4}{2}},\ 4 \le t_2 < t_3 \le 2n,
\end{aligned} \tag{E.13}
$$

and

$$
\begin{aligned}
P\left(Z_{(3)}=t_3\right) &= \sum_{t_2=4}^{2n-1} P\left(Z_{(2)}=t_2, Z_{(3)}=t_3\right) \\
&= \frac{3\left(t_3-1\right)\left(t_3-2\right)\left(t_3-3\right)\left(t_3-4\right)\left(t_3-5\right)}{4\binom{2n}{2}\binom{2n-2}{2}\binom{2n-4}{2}},\quad 6 \le t_3 \le 2n.
\end{aligned} \tag{E.14}
$$

Therefore,

$$
\begin{aligned}
P\left(Z_{(1)}=t_1 \middle| Z_{(2)}=t_2, Z_{(3)}=t_3\right) &= \frac{P\left(Z_{(1)}=t_1, Z_{(2)}=t_2, Z_{(3)}=t_3\right)}{P\left(Z_{(2)}=t_2, Z_{(3)}=t_3\right)} \\
&= \frac{2\left(t_1-1\right)}{\left(t_2-1\right)\left(t_2-2\right)},\quad 2 \le t_1 < t_2 < t_3 \le 2n,
\end{aligned} \tag{E.15}
$$

and

$$
\begin{aligned}
P\left(Z_{(2)}=t_2 \middle| Z_{(3)}=t_3\right) &= \frac{P\left(Z_{(2)}=t_2, Z_{(3)}=t_3\right)}{P\left(Z_{(3)}=t_3\right)} \\
&= \frac{4\left(t_2-1\right)\left(t_2-2\right)\left(t_2-3\right)}{\left(t_3-1\right)\left(t_3-2\right)\left(t_3-3\right)\left(t_3-4\right)},\quad 4 \le t_2 < t_3 \le 2n.
\end{aligned} \tag{E.16}
$$

Now compute conditional expected values

$$
\begin{aligned}
E\left[Z_{(1)} \middle| Z_{(2)}=t_2, Z_{(3)}=t_3\right] &= \sum_{t_1=2}^{t_2-1} t_1 P\left(Z_{(1)}=t_1 \middle| Z_{(2)}=t_2, Z_{(3)}=t_3\right) \\
&= \sum_{t_1=2}^{t_2-1} \frac{2t_1\left(t_1-1\right)}{\left(t_2-1\right)\left(t_2-2\right)} \\
&= \frac{2t_2}{3},
\end{aligned} \tag{E.17}
$$

and

$$E\left[Z_{(2)}E\left[Z_{(1)}\middle|Z_{(2)},Z_{(3)}\right]\middle|Z_{(3)}=t_3\right]=E\left[\frac{2Z_{(2)}^2}{3}\middle|Z_{(3)}=t_3\right]$$

(E.18)
$$=\sum_{t_2=4}^{t_3-1}\frac{2t_2^2}{3}\frac{4\left(t_2-1\right)\left(t_2-2\right)\left(t_2-3\right)}{\left(t_3-1\right)\left(t_3-2\right)\left(t_3-3\right)\left(t_3-4\right)}$$
$$=\frac{4t_3\left(5t_3-1\right)}{45}.$$

Finally,

$$E\left[Y_1Y_2Y_3\right]=E\left[Z_{(3)}E\left[Z_{(2)}E\left[Z_{(1)}\middle|Z_{(2)},Z_{(3)}\right]\middle|Z_{(3)}\right]\right]=E\left[\frac{4Z_{(3)}^2\left(5Z_{(3)}-1\right)}{45}\right]$$

(E.19)
$$=\sum_{t_3=6}^{2n}\frac{4t_3^2\left(5t_3-1\right)}{45}\frac{3\left(t_3-1\right)\left(t_3-2\right)\left(t_3-3\right)\left(t_3-4\right)\left(t_3-5\right)}{4\binom{2n}{2}\binom{2n-2}{2}\binom{2n-4}{2}}$$
$$=\frac{16\left(2n+1\right)\left(70n^2+49n+6\right)}{945}.$$

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX B: QUANTILE TABLES

A. APPROXIMATE CRITICAL VALUES FOR T_N

α	$N = 6$	8	10	12	14	16	18	20	22
0.001	-	-	-	43	58	76	97	120	145
0.005	-	-	31	44	60	79	99	123	149
0.01	-	-	31	45	61	80	101	125	151
0.025	-	21	32	46	63	81	103	127	154
0.05	-	21	33	47	64	83	105	129	156
0.1	13	22	34	48	65	85	107	132	159

α	24	26	28	30	32	34	36	38	40
0.001	173	204	237	272	310	351	394	440	489
0.005	178	209	242	279	317	359	403	449	498
0.01	180	211	245	282	321	363	407	454	503
0.025	183	215	249	286	326	368	413	460	510
0.05	186	218	253	290	330	373	418	466	516
0.1	189	222	257	295	335	378	424	472	523

α	60	80	100	120	140	160	180	200	220
0.001	1113	1995	3137	4537	6199	8121	10304	12749	15455
0.005	1131	2023	3175	4588	6262	8199	10397	12857	15581
0.01	1140	2036	3194	4612	6293	8236	10442	12910	15641
0.025	1152	2056	3221	4648	6338	8292	10508	12987	15731
0.05	1163	2073	3244	4679	6377	8339	10564	13054	15807
0.1	1176	2092	3272	4715	6422	8394	10630	13130	15896

α	240	260	280	300	320	340	360	380	400
0.001	18424	21655	25148	28903	32922	37203	41747	46554	51624
0.005	18567	21816	25328	29103	33142	37444	42009	46838	51931
0.01	18636	21894	25415	29200	33248	37560	42136	46976	52080
0.025	18737	22008	25543	29342	33404	37732	42323	47179	52299
0.05	18825	22107	25653	29464	33539	37879	42483	47353	52487
0.1	18925	22220	25780	29604	33694	38049	42668	47553	52703

Table 19. Estimated critical values for T_N.

Critical regions correspond to values of T_N strictly less than the appropriate quantile; N is sample size and α is significance level. Values for N = 6, 8, and 10 are exact; values for $N > 10$ are approximations by Edgeworth expansion using (3.31). A dash entry ("-") means that the significance level of interest cannot be attained for that sample size.

B. APPROXIMATE CRITICAL VALUES FOR K_N

$N = 20$

α	d = 1	2	3	4	5	10	20	50
0.001	2.57 [.04]	2.40 [.03]	2.27 [.04]	2.17 [.02]	2.10 [.02]	1.93 [.02]	1.88 [.02]	1.78 [.02]
0.005	1.96 [.02]	1.88 [.02]	1.81 [.02]	1.76 [.01]	1.72 [.02]	1.60 [.02]	1.56 [.01]	1.52 [<.01]
0.01	1.72 [.02]	1.66 [.01]	1.60 [.01]	1.56 [.01]	1.53 [.02]	1.46 [.01]	1.43 [<.01]	1.38 [.01]
0.025	1.38 [<.01]	1.35 [.01]	1.33 [.02]	1.31 [.01]	1.29 [.02]	1.24 [<.01]	1.22 [.02]	1.21 [<.01]
0.05	1.13 [<.01]	1.12 [.02]	1.10 [.01]	1.10 [<.01]	1.09 [.01]	1.07 [<.01]	1.07 [.01]	1.03 [.02]
0.1	0.90 [<.01]	0.90 [<.01]	0.89 [<.01]	0.89 [.01]	0.89 [.01]	0.88 [.01]	0.88 [.02]	0.86 [.01]

$N = 40$

α	d = 1	2	3	4	5	10	20	50
0.001	2.77 [.04]	2.57 [.04]	2.42 [.03]	2.31 [.04]	2.22 [.03]	2.02 [.02]	1.95 [.02]	1.84 [.02]
0.005	2.11 [.02]	1.99 [.02]	1.90 [.01]	1.83 [.02]	1.78 [.01]	1.66 [.01]	1.62 [.01]	1.56 [.01]
0.01	1.83 [.01]	1.74 [.01]	1.68 [.01]	1.63 [.01]	1.59 [.01]	1.50 [.01]	1.47 [.01]	1.43 [.01]
0.025	1.47 [.01]	1.42 [.01]	1.38 [.01]	1.35 [.01]	1.33 [.01]	1.28 [.01]	1.26 [.01]	1.24 [<.01]
0.05	1.20 [.01]	1.17 [.01]	1.15 [<.01]	1.14 [<.01]	1.13 [<.01]	1.10 [<.01]	1.09 [<.01]	1.08 [<.01]
0.1	0.93 [<.01]	0.92 [<.01]	0.92 [<.01]	0.91 [<.01]	0.91 [<.01]	0.91 [<.01]	0.91 [<.01]	0.90 [<.01]

$N = 60$

α	d = 1	2	3	4	5	10	20	50
0.001	2.80 [.04]	2.61 [.05]	2.46 [.04]	2.35 [.03]	2.26 [.03]	2.05 [.02]	1.97 [.02]	1.86 [.02]
0.005	2.15 [.02]	2.02 [.02]	1.93 [.02]	1.86 [.01]	1.80 [.01]	1.68 [.01]	1.64 [.01]	1.57 [.01]
0.01	1.85 [.01]	1.76 [.01]	1.70 [.01]	1.65 [.01]	1.62 [.01]	1.53 [.01]	1.50 [.01]	1.44 [.01]
0.025	1.47 [.01]	1.43 [.01]	1.39 [.01]	1.37 [.01]	1.35 [.01]	1.30 [.01]	1.28 [<.01]	1.25 [<.01]

0.05	1.20 [.01]	1.18 [.01]	1.16 [<.01]	1.15 [<.01]	1.14 [<.01]	1.11 [<.01]	1.10 [<.01]	1.09 [<.01]
0.1	0.94 [<.01]	0.93 [<.01]	0.93 [<.01]	0.93 [<.01]	0.92 [<.01]	0.92 [<.01]	0.92 [<.01]	0.92 [<.01]

$N \geq 80$								
α	d = 1	2	3	4	5	10	20	50
0.001	2.88 [.06]	2.67 [.03]	2.50 [.03]	2.38 [.03]	2.29 [.03]	2.06 [.02]	1.98 [.02]	1.88 [.02]
0.005	2.15 [.02]	2.04 [.02]	1.96 [.01]	1.89 [.01]	1.84 [.01]	1.70 [.01]	1.65 [.01]	1.59 [.01]
0.01	1.86 [.01]	1.78 [.01]	1.72 [.01]	1.67 [.01]	1.63 [.01]	1.54 [.01]	1.50 [.01]	1.45 [.01]
0.025	1.49 [.01]	1.45 [.01]	1.41 [.01]	1.38 [.01]	1.36 [.01]	1.31 [.01]	1.29 [.01]	1.26 [<.01]
0.05	1.21 [.01]	1.19 [.01]	1.18 [.01]	1.16 [<.01]	1.15 [<.01]	1.13 [<.01]	1.11 [<.01]	1.10 [<.01]
0.1	0.95 [<.01]	0.94 [<.01]	0.94 [<.01]	0.94 [<.01]	0.93 [<.01]	0.93 [<.01]	0.93 [<.01]	0.92 [<.01]

Table 20. Approximate critical values for K_N.

N is sample size, d is dimension, and α is significance level. Critical values are listed with associated standard error in square brackets. Critical regions correspond to values of K_N strictly greater than the appropriate quantile. Critical values are computed by 100,000 simulations for each case of sample size and dimension using uniformly distributed data and matching with respect to Euclidean distance. Standard error for quantiles is determined by the Maritz-Jarrett method (Maritz and Jarrett, 1978). Simulation suggests that these critical value approximations are independent of underlying distribution and cost function.

Interpolate to find critical values for N, d, or α not provided in the table. Use $d = 50$ to approximate critical values for $d > 50$. For sample size or dimension far outside the bounds of these tables, critical values ought to be approximated by simulation for the case at hand.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX C: R FUNCTIONS FOR CRITICAL ENVELOPES

In this appendix, we provide R (2005) language code to compute critical envelopes for the Non-Bipartite Accumulated Pairs (NAP) and Bipartite Accumulated Pairs (BAP) tests.

A. CRITICAL ENVELOPE FOR THE NAP TEST

```
function(alpha, n){
#  alpha = non-simultaneous alpha value (rejection for
#     exceeding a critical threshold)
#  n = number of pairs (N = 2*n is total sample size)
#  Returned is a list of critical boundary values, and the
#     probability of violating at least one of them.  Boundary
#     values themselves are not critical (e.g. reject the null
#     if any value is strictly greater than the boundary value).
n1 <- n - 1
N <- 2 * n
N1 <- N - 1
qvec <- numeric(N1)
for(k in 2:N1) {
      rmin <- max(c(0, k - n))
      rvec <- rmin:floor(k/2)
      cvec <- cumsum(choose(n, k - rvec) * choose(k - rvec, rvec)
           * 2^(k-2 * rvec))/choose(N, k)
      qvec[k] <- which(cvec > (1-alpha - 1e-010))[1] + rmin-1
}
qvec[1] <- 1
kstar <- max(which(qvec < 1:N1))
a <- rep(0, n)
a[1:2] <- 1
qv <- qvec[2]
for(k in 3:kstar) {
      a0 <- a * ((0:n1) <= qv)
      qv <- qvec[k]
      qv1 <- qv + 1
      if (qv > 0) a[2:qv1] <- a0[2:qv1] - (2 * (diff(a0[1:qv1]) *
           (1:qv)))/k
}
rvec <- max(c(0, kstar - n)):qv
cvec <- (choose(n, kstar - rvec) * choose(kstar - rvec, rvec) *
      2^(kstar - 2 * rvec))/choose(N, kstar)
alpha.sim <- 1 - sum(a[rvec + 1] * cvec)
return(list(kseq <- 2:N1, envelope = qvec[-1], alpha.sim =
      alpha.sim))
}
```

B. CRITICAL ENVELOPE FOR THE BAP TEST

```
function(alpha, m, n){
#  alpha = non-simultaneous alpha value (rejection for
#     exceeding a critical threshold)
#  m = number of control points
#  n = number of test points (must have n > m)
#  Returned is a list of critical boundary values, and the
#     probability of violating at least one of them.  Boundary
#     values themselves are not critical (e.g. reject the null
#     if any value is strictly greater than the boundary value).
if(n <= m) {
     cat("*** Invalid arguments ***", "\n")
     return()
}
qvec <- qhyper(1 - alpha, m, n - m, 1:(n - 1))
sq <- which(diff(c(0, qvec)) < 1e-010)
pvec <- (((m - qvec[sq])/(n - sq + 1)) * dhyper(qvec[sq], m, n -
     m, sq))/phyper(qvec[sq], m, n - m, sq)
return(list(envelope = qvec, alpha.sim = 1 - prod(1 - pvec)))
}
```

APPENDIX D: EXAMPLE DATA FOR FIGURES 1–7

x1	x2
0.8057	0.2209
-1.3556	-1.0061
0.1209	-0.4531
-0.2222	1.3995
0.5717	-0.4620
-0.3001	0.0327
1.1343	0.7988
-0.1794	0.8968
-1.4671	0.1379
1.3953	-1.6191
0.4408	-1.6466
0.5654	0.4287
-0.6936	-0.7372
0.8339	0.5649
-2.2374	-1.3842
1.0976	0.4603
-0.0016	0.6294
-1.6146	0.3798
-1.2287	-1.0133
0.2074	-0.3472

Table 21. Example data for Figures 1–7

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF REFERENCES

Ahuja, R., Magnanti, T., Orlin, J. (1993). *Network Flows*, Prentice Hall, Inc., New Jersey.

Bhattacharya, P. and Frierson, D., Jr. (1981). "A Nonparametric Control Chart for Detecting Small Disorders," *The Annals of Statistics*, Vol. 9, pp. 544–554.

Ball, M. and Derigs, U. (1983). "An Analysis of Alternative Strategies for Implementing Matching Algorithms," *Networks*, Vol. 13, No. 4, pp. 517–549.

Baringhaus, L. and Franz, C. (2004). "On a New Multivariate Two-Sample Test," *Journal of Multivariate Analysis*, Vol. 88, pp. 190–206.

Barnett, V. (1974). "The Ordering of Multivariate Data," *Journal of the Royal Statistical Society Series A (General)*, Vol. 139, No. 3, 318–355.

Basseville, M. and Nikiforov, I. (1993). *Detection of Abrupt Change – Theory and Application*, PTR Prentice-Hall, Inc., Englewood Cliffs, NJ.

Billingsley, P. (1979). *Probability and Measure*, John Wiley & Sons, New York.

Bickel. P. (1969). "A Distribution Free Version of the Smirnov Two Sample Test in the p-Variate Case," *The Annals of Mathematical Statistics*, Vol. 40, No. 1, pp. 33–43.

Brodsky, B. and Darkhovsky, B.S. (1993). *Nonparametric Methods in Change-Point Problems*, pp. 11–24, Kluwer Academic Publishers, Norwell, MA.

Chartrand, G. and Zhang, P. (2005). *Introduction to Graph Theory*, pp. 9, 184, McGraw-Hill, New York, NY, 2005.

Chaudhuri, P. (June 1996). "On a Geometric Notion of Quantiles for Multivariate Data," *Journal of the American Statistical Association*, Vol. 91, No. 434, pp. 862–872.

Choi, K. and Marden, J. (December 1997). "An Approach to Multivariate Rank Tests in Multivariate Analysis of Variance," *Journal of the American Statistical Association*, Vol. 92, No. 440, pp. 1581–1590.

Conover, W. (1999). *Practical Nonparametric Statistics, Third Edition*, pp. 271–288, John Wiley & Sons, New York.

Cook, W. and Rohe A. (February 1999). "Computing minimum-weight perfect matchings," *INFORMS Journal on Computing*, Vol. 11, No. 2, pp. 138–148.

Csörgő, M. and Horváth, L. (1997). *Limit Theorems in Change-Point Analysis*, pp. 7–21, John Wiley & Sons, Inc., New York, NY.

Davies, R. (1997). “Hypothesis Testing When a Nuisance Parameter Is Present Only Under the Alternatives,” *Biometrika*, Vol. 74, No. 1, pp. 33–43.

Derigs, U. (1988). “Solving Non-Bipartite Matching Problems via Shortest Path Techniques,” *Annals of Operations Research,* Vol. 13, pp. 225–261.

Deshayes, J. and Picard, D. (1986). “Off-Line Statistical Analysis of Change-Point Models Using Nonparametric and Likelihood Methods,” *Lecture Notes in Control and Information Sciences*, Vol. 77, pp. 103–168.

Edmonds, J. (1965). “Maximum Matching and a Polyhedron with 0,1-Vertices,” *Journal of Research of the National Bureau of Standards*, Vol. 69B, pp. 125–130.

Fricker, R. and Chang, J. (2009). “The Repeated Two-Sample Rank (RTR) Procedure: A Nonparametric Multivariate Individuals Control Chart,” pre-print dated June 9.

Friedman, J. and Rafsky, L. (1979). “Multivariate Generalizations of the Wald-Wolfowitz and Smirnov Two-Sample Tests,” *The Annals of Statistics,* Vol. 7, No. 4, pp. 697–717.

Fristedt, B. and Gray, L. (1997). *A Modern Approach to Probability Theory*, Birkhäuser, Boston.

Gabow, H.)1973). “Implementation of Algorithms for Maximum Matching on Non-Bipartite Graphs,” Ph.D. Thesis, Computer Science Department, Stanford University.

Gabow, H., Galil, Z., Spencer, T. (1989). “Efficient Implementation of Graph Algorithms Using Contraction,” *Journal of the Association for Computing Machinery*, Vol. 36, No. 3, pp. 540–572.

Galil, Z., Micali, S., Gabow, H. (1986). “An O(*EV* log *V*) Algorithm For Finding A Maximal Weighted Matching In General Graphs,” *SIAM Journal on Computing*, Vol. 15, pp. 120–130.

Girshick, M. and Rubin, H. (March 1952). “A Bayes Approach to a Quality Control Model,” *The Annals of Mathematical Statistics*, Vol. 23, No. 1, pp. 114–125.

Gordon, L. and Pollock, M. (1995). “A Robust Surveillance Scheme for Stochastically Ordered Alternatives,” *The Annals of Statistics*, Vol. 23, pp. 1350–1375.

Greevy, R., Lu, B., Silber, J., Rosenbaum, P. (2004). "Optimal Matching before Randomization," *Biostatistics*, Vol. 5, pp. 263–275.

Hansen, B. (1996). "Inference When a Nuisance Parameter Is Not Identified Under the Null Hypothesis," *Econometrica*, Vol. 64, No. 2, pp. 413–430.

Hastie, T., Tibshirani, R., Freidman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, pp. 501–520, Springer, New York, NY.

Henze, N. and Penrose, M. (1999). "On the Multivariate Runs Test," *The Annals of Statistics*, Vol. 27, No. 1, pp. 290–298.

Hinkley, D. (1969). "Inference about the Intersection in Two-Phase Regression," *Biometrika*, Vol. 56, No. 3, pp. 495–504.

Hodges, J. (1955)."A Bivariate Sign Test," *The Annals of Mathematical Statistics*, Vol. 26, pp. 523–527.

Hotelling, H. (1931). "The Generalization of Student's Ratio," *The Annals of Mathematical Statistics*, Vol. 2, pp. 360–37.

Jonker, R. and Volgenant, A. (1987). "A shortest augmenting path algorithm for dense and sparse linear assignment problems," *Computing*, Vol. 39, pp. 325–340.

Kolmogorov, V. (2009). "Blossom V: A New Implementation of a Minimum Cost Perfect Matching Algorithm," *Mathematical Programming Computation,* Vol. 1, No. 1, pp. 43–67, 2009. Source code in C++ programming language publicly available at http://www.cs.ucl.ac.uk/staff/V.Kolmogorov/software.html#BLOSSOM5 (Retrieved June 24, 2009).

Levedahl, M. (2000). "Performance Comparison of 2-D Assignment Algorithms for Assigning Truth Objects to Measured Tracks," *Proceedings of SPIE – The International Society for Optical Engineering,* Vol. 4048, pp. 380–389.

Li, J. and Liu, R. (2004). "New Nonparametric Tests of Multivariate Locations and Scales Using Data Depth," *Statistical Science*, Vol. 19, No. 4, pp. 686–696.

Liu, R., Parelius, J.M., Singh, K. (1999). "Multivariate analysis by data depth: descriptive statistics, graphics and inference (with discussion and a rejoinder by Liu and Singh)," *The Annals of Statistics*, Vol. 27, No. 3, pp. 783–840, June.

Lowry, C., Woodall, W., Champ, C., Rigdon, S., (1992). "A Multivariate Exponentially Weighted Moving Average Control Chart," *Technometrics*, Vol. 34, pp. 46–53.

Lu, B. and Rosenbaum, P.R. (2004). “An Algorithm for Ranking all the Assignments in Order of Increasing Cost,” *Journal of Computational and Graphical Statistics,* Vol. 13, No. 2, pp. 422–434.

Lu, B., Zanutto, E., Hornik, R., Rosenbaum, P. (2001). “Matching with Doses in an Observational Study of a Media Campaign against Drug Abuse,” *Journal of the American Statistical Association*, Vol. 96, pp. 1245–1253.

Mann, H. and Whitney, D. (1947). “On A Test of Whether One of Two Random Variables Is Stochastically Larger Than the Other,” *The Annals of Mathematical Statistics*, Vol. 18, No. 1, pp. 50–60.

Maple 10.04, (2006). Waterloo Maple Inc., Waterloo, Ontario, Canada.

Maritz, J. and Jarrett, R. (1978). “A Note on Estimating the Variance of the Sample Median,” *Journal of the American Statistical Association*, Vol. 73, No. 361, pp. 194–196.

MATLAB® 7.6.0.324 (R2008a). (2008). The Mathworks, Natick MA.

McKane, B. and Albert, A. (2008). “Distance Functions for Categorical and Mixed Variables,” *Pattern Recognition Letters*, Vol. 29, No. 7, pp. 986–993, May.

Mehlhorn K. and Schäfer, G. (2002). “Implementation of O(*nm*log*n*) Weighted Matchings in General Graphs: The Power of Data Structures,” *Journal of Experimental Algorithmics*, Vol. 7, No. 4.

Möttönen, J. and Oja, H. (1995). “Multivariate Spatial Sign and Rank Methods,” *Nonparametric Statistics*, Vol. 13, No. 5, pp. 201–213.

Murty, K. (1968). “Optimal Pair Matching with Two Control Groups,” *Operations Research*, Vol. 16, No. 3, pp. 682–687.

Osorio, F. and Galea, M. (2005). “Detection of a Change point in Student-t Linear Regression Models,” *Statistical Papers*, Vol. 45, pp. 31–48.

Page, E. (1954). “Continuous Inspection Schemes,” *Biometrika*, Vol. 41, No. 1/2, pp. 100–115, March.

Prabhu, S. and Runger, G. (1997). “Designing A Multivariate EWMA Control Chart,” *Journal of Quality Technology*, Vol. 29, No. 1, pp. 8–15.

Præstgaard, J. (1995). “Permutation and Bootstrap Kolmogorov-Smirnov Tests for the Equality of Two Distributions,” *Scandinavian Journal of Statistics,* Vol. 22, pp. 305–322.

Qui, P. and Hawkins, D. (2003). "A Nonparametric Multivariate Cumulative Sum Procedure for Detecting Shifts in All Directions," *The Statistician,* Vol. 52, pp. 151–164.

R Development Core Team, R: A language and environment for statistical computing, R Foundation for Statistical Computing, (2005). Vienna, Austria, ISBN 3-900051-07-0, Retrieved August 6, 2009, from http://www.R-project.org.

Rinott, Y. and Rotar, V. (2000). "Normal approximations by Stein's method," *Decisions in Economics and Finance,* Vol. 23, pp. 15–29.

Roberts, S. (1959). "Control Chart Tests Based on Geometric Moving Averages," *Technometrics,* Vol. 1, pp. 239–250.

Rosenbaum, P. (2005). "An Exact Distribution-Free Test Comparing Two Multivariate Distributions Based on Adjacency," *Journal of the Royal Statistical Society Series B*, Vol. 67, No. 4, pp. 515–530.

Ross, K. (1981). *Elementary Analysis: The Theory of Calculus*, p. 80, Springer Science & Business, McGraw-Hill, New York, NY.

Runger, G. and Testik, M. (2004). "Multivariate Extensions to Cumulative Sum Control Charts," *Quality and Reliability Engineering International*, Vol. 20, pp. 587–606.

Rushton, S. (1950). "On a Sequential t-Test," *Biometrika*, Vol. 37, No. 3/4, pp. 326–333, December.

Shiryaev, A. (1963). "On Optimum Methods in Quickest Detection Problems," *Theory of Probability and Its Applications,* Vol. 8, No. 1, pp. 22–46.

S-PLUS® Version 7.0. (2005). Insightful Corporation, Seattle, WA. Retrieved August 6, 2009, from http://www.insightful.com.

Stein, C. (1972). "A bound for the error in the normal approximation to the distribution of a sum of dependent random variables," *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 2, pp. 583–602.

Stein, C. (1986). "Approximate Computation of Expectations," *Institute of Mathematical Statistics*, Hayward, CA.

Stoumbos, Z., Reynolds, M., Jr., Ryan, T., Woodall, W. (2000). "The state of statistical process control as we proceed into the 21st century," *Journal of the American Statistical Association*, Vol. 95, No. 451, pp. 992–998.

Tanis, E. and Hogg, R. (2008). *A Brief Course in Mathematical Statistics*, pp. 165–168, Prentice Hall, Inc., Upper Saddle River, NY, 2008.

Tukey, J. (1975). “Mathematics and Picturing Data,” *Proceedings of the 1975 International Congress of Mathematics*, Vol. 2, pp. 523–531.

Wallace, D. (1958). “Approximations to Distributions,” *The Annals of Mathematical Statistics*, Vol. 29, No. 3, pp. 635–654.

Wald, N. and Wolfowitz, J. (1940). “On a Test Whether Two Samples are from the Same Population,” *The Annals of Mathematical Statistics*, Vol. 11, No. 2, pp. 147–162.

Wang, N. and Raftery, A. (2002). “Nearest Neighbor Variance Estimation (NNVE): Robust Covariance estimation via Nearest Neighbor Cleaning,” *Journal of the American Statistical Association*, Vol. 97, No. 460, pp. 994–1019, December.

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
Ft. Belvoir, Virginia

2. Dudley Knox Library
Naval Postgraduate School
Monterey, California

3. Associate Professor Robert Koyak
Naval Postgraduate School
Monterey, California

4. Associate Professor Kyle Lin
Naval Postgraduate School
Monterey, California

5. Associate Professor Craig Rasmussen
Naval Postgraduate School
Monterey, California

6. Associate Professor Javier Salmeron
Naval Postgraduate School
Monterey, California

7. Associate Professor Lyn Whitaker
Naval Postgraduate School
Monterey, California

8. Associate Professor Samuel Buttrey
Naval Postgraduate School
Monterey, California

9. Dr. Eric Bechhoefer
Goodrich Fuels and Utility Systems
Vergennes, Vermont

10. Commander David Ruth, USN
Military Professor, Department of Mathematics
United States Naval Academy
Annapolis, Maryland