

The Pennsylvania State University
The Graduate School
Department of Statistics

SIGN-SCORED METHODS FOR TESTING ORDER RESTRICTED MODELS
IN THE C-SAMPLE DESIGN

A Thesis in
Statistics

by

Scott Reese Preston

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

December 1993

We approve the thesis of Scott Reese Preston.

Date of Signature

Thomas A. Ryan
Associate Professor of Statistics
Thesis Adviser
Chair of Committee

Steven F. Arnold
Associate Professor of Statistics

James L. Rosenberger
Professor of Statistics
Head of the Department of Statistics

Jesse L. Barlow
Professor of Computer Science
and Engineering

Abstract

Hypothesis testing in the c -sample model is generally accomplished using analysis of variance techniques. Such techniques are inappropriate when an ordering of the c populations is enforced a priori. Order restricted methods are effective tools for data analysis in such situations. The primary objective of this dissertation is to establish distributional properties of sign-scored methods for hypothesis testing when order restrictions are assumed. Sign-based methods are desirable because of robustness considerations—the methods are quite resistant to relatively large amounts of outlying data.

Two sign-based approaches are considered for testing an order restricted hypothesis specifying that response location increases with sample index. The first approach is a generalization of the two-sample sign-based procedure, relying on linear combinations of one-sided, two-sample test statistics to test for trend. Distributional results for such statistics are developed in Chapters 1 and 2.

In Chapter 3 the efficiency of the sign-based methods with respect to other procedures based on linear combinations of two-sample statistics is derived. Assuming perfect information regarding the nature of the response curve, it is shown that the linear combination can be weighted to optimize the efficiency of the test. These optimal weightings extend to a broader class of statistics based on linear combinations of rank-based two-sample statistics.

A second sign-based approach for testing an order restricted model is derived as an analogue to the likelihood ratio test for the normal means model. Distributional properties and efficiencies for this test statistic are examined in Chapter 4. Numerical comparisons between the two methods examined indicate that, when

the precise nature of the response curve can be anticipated, the optimally weighted linear combination of two-sample statistics is superior. It is also demonstrated that the optimally weighted tests are preferable when the true and anticipated response curves do not drastically differ. In practice a linear response is often hypothesized; the optimally weighted test against such a response performs very well provided that the deviation from this model is not too large.

Table of Contents

List of Tables	viii
List of Figures	ix
Acknowledgments	x
Chapter 1 Introduction and Preliminary Results	1
1.1 Overview	1
1.1.1 Order restricted methods	1
1.1.2 Mood's median test	2
1.1.3 Linear contrast methods	3
1.1.4 Likelihood ratio tests and distribution-free variations	5
1.1.5 Outline	6
1.2 Assumptions and notation	7
1.3 An alternate expression for the Mood median statistic	8
1.4 Multivariate limiting normality	11
1.5 Application	14
1.6 Remark	15
Chapter 2 Limiting Distribution Under the Alternate Hypothesis	16
2.1 Preliminaries	16
2.2 An alternate expression for the Mood median statistic	22
2.3 Multivariate limiting normality	29

Chapter 3	Asymptotic Relative Efficiency	33
3.1	Asymptotic distribution under a sequence of shift alternatives	34
3.2	Asymptotic relative efficiency	37
3.3	Equivalent statistics	42
3.4	Optimal weighting coefficients	45
3.5	Examples	48
3.6	Remarks	51
Chapter 4	Comparison of Procedures	54
4.1	Efficiency comparisons	54
4.2	Sign-scored $\bar{\chi}^2$ test	65
4.3	Simulations	69
4.3.1	Design	69
4.3.2	Results	72
Chapter 5	Resistance to Rejection of the One and Two-Sample Median Tests	78
5.1	Motivation	78
5.1.1	Expected resistance to rejection	79
5.1.2	Extensions	81
5.2	Two-sample setting, single sample contamination schemes	82
5.3	Optimal contamination scheme	90
5.4	Bounding results	91
5.5	Limiting distribution of the single sample scheme	96
5.6	A partition	97
5.7	Equivalency of the optimal and smaller sample schemes	100

5.7.1	Unequal sample sizes	101
5.7.2	Equal sample sizes	102
5.7.3	Equivalence	103
5.8	Distance to rejection	103
5.9	Remark	105
Chapter 6 Resistance To Rejection of the Wilcoxon Signed Rank Test		106
6.1	Preliminaries	107
6.2	Contamination points	108
6.3	Limiting distribution	109
6.4	Remarks	114
Chapter 7 Summary		116
Bibliography		118

List of Tables

4.1	Asymptotic Relative Efficiencies, Equal Sample Sizes	61
4.2	Asymptotic Relative Efficiencies, Descending Sample Sizes	62
4.3	Asymptotic Relative Efficiencies, Lightly Control Dominated	63
4.4	Asymptotic Relative Efficiencies, Heavily Control Dominated	64
4.5	Simulated Power, Equal Sample Sizes	74
4.6	Simulated Power, Descending Sample Sizes	75
4.7	Simulated Power, Lightly Control Dominated	76
4.8	Simulated Power, Heavily Control Dominated	77
5.1	Expected Resistance to Rejection of the One-Sided One-Sample Sign Test	81
5.2	Expected Resistance to Rejection of Mood's Test: Smaller Sample Contamination	89
6.1	Simulated and Approximate Values for the Resistance to Rejection of the Wilcoxon Test	115

List of Figures

1.1	Structure of the A Matrix Under H_0	12
2.1	Structure of the A Matrix Under H_1	30
3.1	Asymptotic Relative Efficiency of Optimally Weighted Tests	50
4.1	Response Curve A: Linear	56
4.2	Response Curve B: Linear With Double Spacing Between the Middle Groups	56
4.3	Response Curve C: Linear With Ten-fold Spacing Between the First Two Groups	57
4.4	Response Curve D: Convex Quadratic	57
4.5	Response Curve E: Convex Exponential	58
4.6	Response Curve F: Partial Sums of Harmonic Series	58
4.7	Response Curve G: Concave Exponential	59
4.8	Response Curve H: Concave Quadratic	59
4.9	Response Curves Used in the Simulation	70

Acknowledgments

A number of fine people have contributed to this effort. Ford Carpenter, Al Reilingh, Miles Preston and Eva Petkova have engaged me in many fruitful conversations regarding both product and process, preparing me to meet the challenge of the new frontier. Regis Serinko deserves my gratitude for his time, patience and advice on a number of technical matters. Thomas P. Hettmansperger provided me with invaluable instruction and supplied a key idea that runs throughout this work.

I am proud to have earned the approval of such a fine committee. The instruction, motivation, support and friendship of Steve Arnold and Jim Rosenberger are greatly appreciated. My adviser, Tom Ryan, deserves special mention, not only for the invaluable advice and insight he has supplied me with, but also for providing me with a clear perspective on my future as a statistician.

I am ever indebted to the members of my family for their concern, interest and patience. My parents, Mr. and Mrs. Malcolm M. Preston, my sisters, Beth, Judi and Cyndy, and my grandmother Mrs. William S. Preston all deserve my thanks.

For the many opportunities and privileges life has offered me, I thank my grandmother, Mrs. Charles H. Reese.

Chapter 1

Introduction and Preliminary Results

This thesis begins by developing sign-scored methods for testing against ordered alternatives in the c -sample design. Considerable work has been done on the topic of order restricted inference; Robertson, Wright, and Dykstra (1988) provide a comprehensive reference. None of this work implicitly or explicitly develops sign-scored methods. Rectifying this situation is the initial focus of this paper.

We first develop a generalization of Mood's two-sample median test; in particular we use a linear combination of one-sided Mood statistics for testing $\eta_i < \eta_j$ where the index set is $1 \leq i < j \leq c$. Following this, the sign-scored analogue of the $\bar{\chi}^2$ statistic developed by Bartholomew (1959) is examined.

1.1 Overview

1.1.1 Order restricted methods

Testing and estimation in the c -sample model are generally accomplished through analysis of variance techniques; the F-test and distribution free analogues to the χ^2 -test are prominent. In many applications it is reasonable to assume more about the nature of the response than inequality of parameters between treatment groups. Testing order restricted models in the c -sample problem is an extension of the 1-sided test in the 2-sample problem.

This work focuses on procedures for testing the "simple order" alternate hypothesis versus the null hypothesis of homogeneity of response over the treatment

groups, i.e.,

$$H_0 : F_1(x) = F_2(x) = \cdots = F_c(x)$$

$$H_1 : F_1(x) \geq F_2(x) \geq \cdots \geq F_c(x) \quad (F_i(x) > F_{i+1}(x), \text{ some } i).$$

There are a number of approaches to testing such hypotheses, including normal theory likelihood ratio tests and linear contrast methods, as well as distribution free variations on these. When the general alternative is dismissed in favor of the simple order alternative, each of these methods is quite effective in increasing procedural power for alternatives in H_1 .

1.1.2 Mood's median test

Mood (1954) is credited with the development of the distribution theory for the sign-scored, 2-sample testing procedure. For the 2-sample design, with X_{il} , $i = 1, 2$, $l = 1, \dots, n_i$, independent and distributed according to some distribution function F_i , Mood's test is based on the statistic

$$M = \#(X_{2l} > \text{median} \{X_{il}; i = 1, 2, l = 1, \dots, n_i\}).$$

Under H_0 , assuming some mild regularity conditions, the exact distribution of M is hypergeometric, M is asymptotically normal. Mood's test has the advantage of being quite robust to outlying data.

Mood (1954) also developed the asymptotic distribution theory of the 2-sample statistic M under a general hypothesis. His work was later extended by Andrews (1954) in deriving limit results regarding the extension of the 2-sample Mood test to the c -sample setting, unrestricted alternative. The technique employed by both Mood and Andrews is to compute the exact joint distribution of the aggregate

sample median and the number of observations from each sample left of this median, “use Stirling’s formula on the factorials, take logarithms, . . .,” Mood (1954). Similar, but separate developments must be taken for even and odd aggregate sample sizes. Such an approach is clearly ill-suited to development of the distribution theory of linear combinations of a number of such (dependent) statistics.

The work of Bahadur (1966) on quantile representation theory post-dates that of Mood and Andrews. The approach taken here applies methods of Bahadur in developing the limiting distribution of linear combinations of 2-sample Mood statistics. The method leads to a significantly more concise treatment of the problem, is applicable to both even and odd sample sizes, and lends itself to an intuitive interpretation.

1.1.3 Linear contrast methods

Contrasts of sample means have long been known to outperform the F -test when testing for trend. The initial work in distribution-free testing for trend in the c -sample model was done by Terpstra (1953) and Jonckheere (1954), who independently derived the distributional properties and cumulant generating function for a statistic based on a linear combination of one-sided Mann-Whitney tests:

$$J = \sum_{i=1}^{c-1} \sum_{j=i+1}^c W_{ij},$$

where W_{ij} is the Mann-Whitney test for comparing the i th and j th samples.

This work was extended by Puri (1965) who studied the statistic which replaces W_{ij} by a Chernoff-Savage statistic (Chernoff & Savage, 1958). Let $\hat{F}_i(x)$

denote the empirical distribution function for the i -th sample. Define

$$\hat{F}_{ij}(x) = \frac{n_i}{n_i + n_j} \hat{F}_i(x) + \frac{n_j}{n_i + n_j} \hat{F}_j(x), \quad (1.1)$$

the combined sample distribution function of the i th and j th samples. Then take

$$W = \sum_{i=1}^{c-1} \sum_{j=i+1}^c n_i n_j w_{ij}, \quad (1.2)$$

where w_{ij} is expressed:

$$w_{ij} = \int_{-\infty}^{+\infty} J_{n_i+n_j} [\hat{F}_{ij}(x)] d(\hat{F}_i(x) - \hat{F}_j(x)).$$

$J_{n_i+n_j}$ is a score function satisfying Assumptions 1 - 4 of Puri (1965). To generate the sign-scored version of this test, for $1 \leq i < j \leq c$, set

$$J_{n_i+n_j} [\nu/(n_i + n_j)] = \begin{cases} 0 & \frac{\nu}{n_i + n_j} \leq \frac{1}{2}, \\ 1 & \frac{\nu}{n_i + n_j} > \frac{1}{2}, \end{cases} \quad \nu = 1, 2, \dots, n_i + n_j. \quad (1.3)$$

Then (1.2) becomes

$$W = \sum_{i=1}^{c-1} \sum_{j=i+1}^c (n_i + n_j) (M_{ij} - n_j/2), \quad (1.4)$$

where M_{ij} is the Mood median statistic for samples i and j .

One of Puri's restrictions on the score function $J_N(\cdot)$ is

$$\left| \frac{d}{du} J(u) \right| \leq K [u(1-u)]^{\delta-3/2},$$

for some K and some $\delta > 0$, where

$$J(u) = \lim_{N \rightarrow \infty} J_N(u), \quad 0 < u < 1.$$

For the score function defined in (1.3), $J(u)$ has a jump discontinuity at $u = 1/2$, hence $J(u)$ is not differentiable at $u = 1/2$. Puri's results do not apply here; nor does a generalization seem to be available. As such, the corresponding results will be developed using different methods.

1.1.4 Likelihood ratio tests and distribution-free variations

The normal theory likelihood ratio test for simple order was developed by Bartholomew (1959):

$$\bar{\chi}^2 = \sigma^{-2} \sum_{i=1}^c n_i (X_i^* - \bar{X})^2,$$

where the X_i^* are the isotonic regression of the (unrestricted) sample means, \bar{X}_i .

Distribution-free analogues to the $\bar{\chi}^2$ test replace observations in the aggregate sample with scores $J_N(u)$, where N is the aggregate sample size and $J_N(\cdot)$ is again a Chernoff-Savage statistic. Let R_{ij} be the rank of observation j in subsample i , then $\bar{S}_i = \sum_{j=1}^{n_i} J_N(R_{ij})/n_i$ is the average score in subsample i and $\bar{J}_N = \sum_{i=1}^N J_N(i)/N$ is the average score over all samples. Define

$$\sigma_{J_N}^2 = \sum_{i=1}^N [J_N(i) - \bar{J}_N]^2 / (N - 1).$$

The vector of average scores within subsamples is given by $\bar{\mathbf{S}} = [\bar{S}_1, \bar{S}_2, \dots, \bar{S}_c]^T$.

Denote the isotonic regression of $\bar{\mathbf{S}}$ by

$$\bar{\mathbf{S}}^* = [\bar{S}_1^*, \bar{S}_2^*, \dots, \bar{S}_c^*]^T.$$

Then

$$\bar{\chi}^2(\bar{\mathbf{S}}) = \frac{\sum_{i=1}^c n_i [\bar{S}_i^* - \bar{J}_N]^2}{\sigma_{J_N}^2}$$

has the limiting distribution of a $\bar{\chi}^2$ random variable. Chacko (1963) developed results for $J_N(i) = i/N$ and equal sample sizes; Shorack (1967) extended them to the unequal sample sizes case. Shiraishi (1982) generalized these methods to allow for score functions satisfying the conditions of Chernoff and Savage.

The sign-scored analogue to these statistics is generated by applying the score function J_N as in (1.3). Take $\hat{\eta}$ to be the aggregate sample median and $[\hat{F}_1^*(\hat{\eta}), \dots, \hat{F}_c^*(\hat{\eta})]^T$ to be the antitonic regression of the vector $[\hat{F}_1(\hat{\eta}), \dots, \hat{F}_c(\hat{\eta})]^T$. Then

$$\bar{\chi}^2(\bar{\mathbf{S}}) = \sum_{i=1}^c 4n_i \left[\hat{F}_i^*(\hat{\eta}) - \frac{1}{2} \right]^2,$$

Again, the cited distribution theory does not cover this case.

1.1.5 Outline

Chapter 1 concludes by developing a re-expression of the Mood statistic. The results are stronger than necessary to derive the asymptotic distribution theory of the statistic W of (1.3). They will be implemented in full in Chapters 5 and 6, where an unrelated topic, the resistance to rejection of common non-parametric testing procedures, will be examined. Chapter 2 develops the distributional theory for W under a general hypothesis. In Chapter 3 asymptotic relative efficiency and related topics are taken up. Having provided all the details in our arguments supporting this alternate approach to Mood statistics, in Chapter 4 we develop results for the sign-scored analogue to the $\bar{\chi}^2$ test and present numeric comparisons of the

various procedures.

1.2 Assumptions and notation

The over-all sample consists of $N = \sum_{i=1}^c n_i$ independent random variables X_{il} , $i = 1, \dots, c$, $l = 1, \dots, n_i$, where the first subscript refers to the subsample and the second subscript indexes observations within a subsample. Under the null hypothesis all of the X_{il} 's have the same, but unknown, cumulative distribution function $F(x)$. Make the following assumptions on $F(\cdot)$:

A1. There exists a unique value η such that $F(\eta) = 1/2$, and

A2. F has non-vanishing and differentiable density $f(\cdot)$ at the point η .

A proportional sampling scheme is assumed throughout; for each $i = 1, \dots, c$, $\lim_{N \rightarrow \infty} n_i/N = \lambda_i$, where $0 < \lambda_i < 1$ and $\sum_{i=1}^c \lambda_i = 1$.

Let $\hat{\eta}_{ij}$ be the sample median for samples i and j combined, uniquely defined by

$$\hat{\eta}_{ij} = \inf\{x : \hat{F}_{ij}(x) \geq 1/2\} = \hat{F}_{ij}^{-1}(1/2).$$

Throughout, we shall make use of a representation of Mood's statistic in terms of the empirical distribution function. Define

$$M_{ij} = n_i \hat{F}_i(\hat{\eta}_{ij}),$$

then M_{ij} is the number of sample i observations not greater than the median of samples i and j .

1.3 An alternate expression for the Mood median statistic

In order to re-express the Mood statistic a result of Bahadur (1966) is applied. Let $n = n_1 + n_2$; without loss of generality the result is proved for $i = 1$, $j = 2$. Proofs of the following two lemmas are given in Serfling (1980), pages 95–98.

Lemma 1.1 *Assume A1 and A2 are satisfied. Then*

$$P \left[|\hat{\eta}_{12} - \eta| \geq \frac{2(\log n)^{1/2}}{f(\eta)n^{1/2}} \text{ i.o.} \right] = 0. \quad (1.5)$$

Lemma 1.2 (Bahadur (1966)) *Assume A1 and A2 are satisfied. For either $k = 1$ or $k = 2$ let $\{a_{n_k}\}$ be a sequence of positive constants, $a_{n_k} \sim Cn_k^{-1/2}(\log n_k)^{1/2}$, where $C > 0$. Set*

$$\hat{H}_k = \sup_{|x| < a_{n_k}} |[F_{n_k}(\eta + x) - F_{n_k}(\eta)] - [F(\eta + x) - F(\eta)]|. \quad (1.6)$$

Then there exists a constant M_k such that

$$P \left[\hat{H}_k \geq M_k n_k^{-3/4} (\log n_k)^{3/4} \text{ i.o.} \right] = 0. \quad (1.7)$$

The following theorem demonstrates that the Mood statistic has an asymptotic representation as a weighted difference of one-sample sign statistics.

Theorem 1.1 *Assume A1 and A2 are satisfied. Then*

$$\left[\hat{F}_1(\hat{\eta}_{12}) - 1/2 \right] = \frac{n_2}{n} \left[\hat{F}_1(\eta) - 1/2 \right] - \frac{n_2}{n} \left[\hat{F}_2(\eta) - 1/2 \right] + R_n, \quad (1.8)$$

where, with probability one, $R_n = O(n^{-3/4}(\log n)^{3/4})$ as $n \rightarrow \infty$.

Proof: Define

$$A_{n_1} = \left[\omega : |\hat{\eta}_{12} - \eta| \geq \frac{2(\log n_1)^{1/2}}{f(\eta) n_1^{1/2}} \right].$$

For n sufficiently large

$$A_{n_1} \subset \left[\omega : |\hat{\eta}_{12} - \eta| \geq \frac{2(\log n)^{1/2}}{f(\eta) n^{1/2}} \right],$$

and Lemma 1.1 implies $P[A_{n_k} \text{ i.o.}] = 0$. Apply Lemma 1.2 with the sequence $\{a_{n_1}\}$ chosen with $a_{n_1} > (2/f(\eta))n^{-1/2}(\log n)^{1/2}$ for all n . Then there exists M_1 such that the sets

$$C_{n_1} = \left[\omega : H_{n_1} \geq M_1 n_1^{-3/4} (\log n_1)^{3/4} \right],$$

satisfy $P[C_{n_1} \text{ i.o.}] = 0$. Further, define

$$B_{n_1} = \left[\omega : \left| [\hat{F}_1(\hat{\eta}) - \hat{F}_1(\eta)] - [F(\hat{\eta}) - F(\eta)] \right| \geq M_1 n_1^{-3/4} (\log n_1)^{3/4} \right].$$

Then

$$B_{n_1} = \left((A_{n_1} \cap B_{n_1}) \cup (A_{n_1}^c \cap B_{n_1}) \right) \subset (A_{n_1} \cup (A_{n_1}^c \cap B_{n_1})) \subset (A_{n_1} \cup C_{n_1}),$$

implying that $\left(\limsup_{n_1} B_{n_1} \right) \subset \left(\limsup_{n_1} (A_{n_1} \cup C_{n_1}) \right)$. Hence

$$\begin{aligned} P \left[\limsup_{n_1} B_{n_1} \right] &\leq P \left[\limsup_{n_1} (A_{n_1} \cup C_{n_1}) \right] \\ &= P \left[\left(\limsup_{n_1} A_{n_1} \right) \cup \left(\limsup_{n_1} C_{n_1} \right) \right] \\ &\leq P \left[\limsup_{n_1} A_{n_1} \right] + P \left[\limsup_{n_1} C_{n_1} \right] = 0. \end{aligned}$$

Rephrasing, with probability one,

$$\left[\hat{F}_1(\hat{\eta}_{12}) - \hat{F}_1(\eta) \right] - [F(\hat{\eta}_{12}) - F(\eta)] = O\left(n_1^{-3/4} (\log n_1)^{3/4}\right), n \rightarrow \infty. \quad (1.9)$$

An identical development shows that, with probability one,

$$\left[\hat{F}_2(\hat{\eta}_{12}) - \hat{F}_2(\eta) \right] - [F(\hat{\eta}_{12}) - F(\eta)] = O\left(n_2^{-3/4} (\log n_2)^{3/4}\right), n \rightarrow \infty. \quad (1.10)$$

As $\hat{F}(\hat{\eta}_{12}) = 1/2 + O(n^{-1})$, use (1.1) to write

$$\hat{F}_2(\hat{\eta}_{12}) = 1/2 - \frac{n_1}{n_2} \hat{F}_1(\hat{\eta}_{12}) + O(n^{-1}). \quad (1.11)$$

Substitute (1.11) into (1.10), subtract the result from (1.9). Algebra then reduces the result to (1.8). ■

The theorem holds for each $1 \leq i < j \leq c$; for our purposes the appropriately scaled version is given by

$$\begin{aligned} N^{-3/2} (n_i + n_j) n_i \left[\hat{F}_i(\hat{\eta}_{ij}) - 1/2 \right] = \\ \left(\frac{n_j}{N} \right) \left(\frac{n_i}{N} \right)^{1/2} \sqrt{n_i} \left[\hat{F}_i(\eta) - 1/2 \right] \\ - \left(\frac{n_i}{N} \right) \left(\frac{n_j}{N} \right)^{1/2} \sqrt{n_j} \left[\hat{F}_j(\eta) - 1/2 \right] + R_{\{n_i, n_j\}}, \end{aligned} \quad (1.12)$$

where, with probability one, $R_{\{n_i, n_j\}} = O(N^{-1/4} (\log N)^{3/4})$, as $N \rightarrow \infty$. In what follows it is sufficient to take $R_{\{n_i, n_j\}} = o(1)$, with probability one, as $N \rightarrow \infty$.

1.4 Multivariate limiting normality

Let $d = c(c-1)/2$ and let \mathbf{V}_N be the d -vector of centered pairwise median statistics:

$$\mathbf{V}_N = \begin{bmatrix} (n_1 + n_2) (M_{12} - n_1/2) \\ (n_1 + n_3) (M_{13} - n_1/2) \\ \vdots \\ (n_{c-1} + n_c) (M_{(c-1)c} - n_{(c-1)}/2) \end{bmatrix}.$$

Adopt the convention of indexing the components of \mathbf{V}_N with double subscripts, $(i, j) : 1 \leq i < j \leq c$. Another representation is given by

$$(\mathbf{V}_N)_{ij} = (n_i + n_j) n_i \left[\hat{F}_i(\hat{\eta}_{ij}) - 1/2 \right].$$

Let \mathbf{S}_N be the c -vector of (centered and scaled) sign statistics,

$$\mathbf{S}_N = \begin{bmatrix} \sqrt{n_1} (\hat{F}_1(\eta) - 1/2) \\ \vdots \\ \sqrt{n_c} (\hat{F}_c(\eta) - 1/2) \end{bmatrix}.$$

From (1.12),

$$N^{-3/2} \mathbf{V}_N = A_N B_N \mathbf{S}_N + \mathbf{R}_N, \tag{1.13}$$

where A_N , B_N and \mathbf{R}_N are as follows.

$$A_N = \begin{bmatrix} \frac{n_2}{N} & -\frac{n_1}{N} & 0 & 0 & 0 \\ \frac{n_3}{N} & 0 & -\frac{n_1}{N} & 0 & 0 \\ \frac{n_4}{N} & 0 & 0 & -\frac{n_1}{N} & 0 \\ \frac{n_5}{N} & 0 & 0 & 0 & -\frac{n_1}{N} \\ 0 & \frac{n_3}{N} & -\frac{n_2}{N} & 0 & 0 \\ 0 & \frac{n_4}{N} & 0 & -\frac{n_2}{N} & 0 \\ 0 & \frac{n_5}{N} & 0 & 0 & -\frac{n_2}{N} \\ 0 & 0 & \frac{n_4}{N} & -\frac{n_3}{N} & 0 \\ 0 & 0 & \frac{n_5}{N} & 0 & -\frac{n_3}{N} \\ 0 & 0 & 0 & \frac{n_5}{N} & -\frac{n_4}{N} \end{bmatrix}.$$

Figure 1.1: Structure of the A Matrix Under H_0

- A_N is a $d \times c$ matrix with components

$$(A_N)_{(ij),k} = \begin{cases} \frac{n_j}{N} & i = k, \\ -\frac{n_i}{N} & j = k, \\ 0 & \text{elsewhere.} \end{cases}$$

The structure of the matrix A_N for the case $c = 5$ is displayed in Figure 1.1.

- B_N is a $c \times c$ diagonal matrix, $(B_N)_{ii} = (n_i/N)^{1/2}$.
- \mathbf{R}_N is a d -vector of terms with (ij) term $o(1)$, with probability one, as $N \rightarrow \infty$.

We now establish the limiting null distribution of the collection of pair-wise Mood statistics.

Theorem 1.2 *Assuming conditions A1 and A2, as $N \rightarrow \infty$ the random vector $N^{-3/2}\mathbf{V}_N$ converges in law to a d -variate normal random vector with mean vector $\mathbf{0}$ and covariance matrix $\Sigma_{d \times d}$ given by*

$$\sigma_{(ij)(km)} = \begin{cases} \frac{\lambda_i \lambda_j (\lambda_i + \lambda_j)}{4} & i = j, k = m, \\ \frac{\lambda_i \lambda_j \lambda_m}{4} & i = k, j \neq m, \\ \frac{\lambda_i \lambda_j \lambda_k}{4} & i \neq k, j = m, \\ -\frac{\lambda_i \lambda_j \lambda_m}{4} & j = k, \\ 0 & i, j, k, m \text{ distinct.} \end{cases} \quad (1.14)$$

Proof: It is well known that the vector \mathbf{S}_N consists of independent components and, as $n_i \rightarrow \infty$, $i = 1, \dots, c$, \mathbf{S}_N converges in law to a c -variate normal random vector with $\mathbf{0}$ mean and covariance matrix $(1/4)I$.

- $A_N \rightarrow A_{d \times c}$

$$(A)_{(ij),k} = \begin{cases} \lambda_j & i = k, \\ -\lambda_i & j = k. \\ 0 & \text{elsewhere.} \end{cases} \quad (1.15)$$

(The structure of the matrix A is displayed for the case $c = 5$ in Figure 1.1, page 12. Replacing the terms n_i/N with λ_i gives A .)

- $B_N \rightarrow B_{c \times c}$ where B is diagonal, $(B)_{ii} = \lambda_i^{1/2}$.
- $\mathbf{R}_N \rightarrow \mathbf{0}$, almost surely.

By the multivariate version of Slutsky's Theorem (Arnold (1990), page 251),

$$N^{-3/2}\mathbf{V}_N = A_N B_N \mathbf{S}_N + \mathbf{R}_N \xrightarrow{\mathcal{L}} \mathbf{X},$$

where \mathbf{X} is d -variate normal with mean vector $AB\mathbf{0} = \mathbf{0}$ and covariance matrix given by $\Sigma = (1/4)ABB^T A^T$. Multiplying the matrices yields the solution given in (1.14). ■

1.5 Application

An immediate implication of Theorem 1.2 is that any linear combination of the components of the random vector $N^{-3/2}\mathbf{V}_N$ will converge in law to a univariate normal. Let \mathbf{b} be a $d \times 1$ vector of coefficients (weights), again indexed with double subscripts, $(\mathbf{b})_{ij} = b_{ij}$, $1 \leq i < j \leq c$. Define

$$V_N(\mathbf{b}) = \mathbf{b}^T \mathbf{V}_N = \sum_{i=1}^{c-1} \sum_{j=i+1}^c b_{ij} (n_i + n_j) \left[M_{ij} - \frac{n_i}{2} \right].$$

Corollary 1.1 *Assume A1 and A2. As $N \rightarrow \infty$,*

$$N^{-3/2}V_N(\mathbf{b}) = N^{-3/2}\mathbf{b}^T \mathbf{V}_N \xrightarrow{\mathcal{L}} Y \sim n(0, \tau^2)$$

where $\tau^2 = \mathbf{b}^T \Sigma \mathbf{b}$.

To test the simple order, assuming the coefficients b_{ij} are equal (without loss of generality $b_{ij} = 1$, $\mathbf{b} = \mathbf{1}$), we have $\tau^2 = (1 - \sum_{i=1}^c \lambda_i^3)/12$. Defining $\sigma_N^2 = (N^3 - \sum_{i=1}^c n_i^3)/12$ gives $\tau/(N^{-3/2}\sigma_N) \rightarrow 1$; as a result $V_N(\mathbf{1})/\sigma_N \xrightarrow{\mathcal{L}} Z \sim n(0, 1)$.

The asymptotic size- α test rejects H_0 for

$$\sum_{i=1}^{c-1} \sum_{j=i+1}^c (n_i + n_j) M_{ij} > \sum_{i=1}^{c-1} \sum_{j=i+1}^c \frac{(n_i + n_j) n_i}{2} + z^\alpha \left[\frac{N^3 - \sum_{i=1}^c n_i^3}{12} \right]^{1/2}.$$

As a special case, consider equal sample sizes (i.e., $n_i = n$, $\lambda_i = c^{-1}$, $i = 1, \dots, c$, and $N = cn$) . Then

$$\sum_{i=1}^{c-1} \sum_{j=i+1}^c M_{ij} > \frac{N(c-1)}{2} + z^\alpha \sqrt{\frac{N(c^2-1)}{48}}$$

is the asymptotic size- α test.

1.6 Remark

Proofs of Lemmas 1.1 and 1.2 can easily be adapted to accommodate any distribution function having density f with finite left and right-sided derivatives at the median η . All results given in this chapter are therefore valid under these relaxed regularity conditions. The double exponential is one model for which the extension is necessary.

Chapter 2

Limiting Distribution Under the Alternate Hypothesis

Under the alternate hypothesis assume X_{il} are independently distributed random variables having distribution functions depending only on i , $X_{il} \sim F_i(x)$, with the $F_i(x)$ stochastically ordered,

$$\forall x : F_1(x) \geq F_2(x) \geq \dots \geq F_c(x).$$

The same sampling scheme used previously is adopted. Notation remains unchanged except where specified. A quantile representation theorem is developed for independent, but not identically distributed random variables. Using this, and a result similar to Lemma 1.2, Mood's statistic is re-expressed in much the same fashion as in Theorem 1.1.

2.1 Preliminaries

In developing the results of this chapter it is necessary to construct sequences of (paired) "population" medians dependent on sample sizes. Take $F(\cdot)$ and $G(\cdot)$ to be stochastically ordered distribution functions: for all x , $F(x) \geq G(x)$. For $0 < y < 1$ define:

$$H(x, y) = yF(x) + (1 - y)G(x) - \frac{1}{2}.$$

Assume that the equation $H(x, y) = 0$ has a solution (x_0, y_0) , and that $F(x)$ and $G(x)$ have continuous, strictly positive densities ($f(x)$ and $g(x)$, respectively) in

some neighborhood of the point x_0 . Therefore $H(x, y)$ has continuous partial derivatives with respect to both x and y in some neighborhood of (x_0, y_0) , and

$$\frac{\partial}{\partial x} H(x, y) |_{(x_0, y_0)} = y_0 f(x_0) + (1 - y_0) g(x_0) > 0.$$

Applying the implicit function theorem, there exists a function $\eta(y)$, having a continuous first derivative in a neighborhood \mathbf{N} of y_0 such that $x = \eta(y)$ is a solution to $H(x, y) = 0$ for y in \mathbf{N} , and such that $x_0 = \eta(y_0)$.

Denote by Ω the class of admissible hypotheses given by the following restrictions.

A1. The F_i are stochastically ordered, with $F_i(x) \geq F_{i+1}(x)$, $-\infty < x < +\infty$,
 $i = 1, \dots, c - 1$.

A2. For each $(i, j) : 1 \leq i \neq j \leq c$ there exists a value η_{ij}° such that

$$\frac{\lambda_i}{\lambda_i + \lambda_j} F_i(\eta_{ij}^\circ) + \frac{\lambda_j}{\lambda_i + \lambda_j} F_j(\eta_{ij}^\circ) = \frac{1}{2}.$$

(For convenience, $\eta_{ij}^\circ = \eta_{ji}^\circ$.)

A3. For each $(i, j) : 1 \leq i \neq j \leq c$, both $F_i(x)$ and $F_j(x)$ have continuous, strictly positive density in some neighborhood of η_{ij}° . As a result, $0 < F_i(\eta_{ij}^\circ) < 1$.

For each $i \neq j$, A1–A3 ensure construction of the function $\eta_{ij}(y)$ as follows:

$$\eta_{ij}(y) : y F_i(\eta_{ij}(y)) + (1 - y) F_j(\eta_{ij}(y)) = \frac{1}{2},$$

where $\eta_{ij}(y)$ has continuous derivative in a neighborhood \mathbf{N}_{ij} of $y = \lambda_i / (\lambda_i + \lambda_j)$.

Since

- For each $i \neq j$, $n_i / (n_i + n_j) \rightarrow \lambda_i / (\lambda_i + \lambda_j)$,

- $\eta_{ij}(y)$ is a continuous function in some neighborhood of $y_0 = \lambda_i/(\lambda_i + \lambda_j)$, and
- $F_i(\cdot)$ is continuous with continuous density $f_i(\cdot)$ in some neighborhood of η_{ij}° ,

there then exists N_0 such that for all $N > N_0$, and each $i \neq j$, the following results hold.

- $n_i/(n_i + n_j)$ is in the domain of the function $\eta_{ij}(y)$, and therefore there exists an explicit solution to

$$\eta_{ij} : \frac{n_i}{n_i + n_j} F_i(\eta_{ij}) + \frac{n_j}{n_i + n_j} F_j(\eta_{ij}) = \frac{1}{2}.$$

The solutions η_{ij} define the paired population medians for sufficiently large values of N . For notational compactness the dependence of the $\eta_{ij}(n_i/(n_i + n_j))$ upon $n_i/(n_i + n_j)$ will be suppressed.

- $0 < \lim_{N \rightarrow \infty} F_i(\eta_{ij}) = F_i(\eta_{ij}^\circ) < 1$.
- $\lim_{N \rightarrow \infty} f_i(\eta_{ij}) = f_i(\eta_{ij}^\circ) > 0$.

Note that η_{ij} may not be well defined for $N \leq N_0$. In such a case it may be defined arbitrarily. In all limit theorems to follow it is assumed that $N > N_0$. If the densities $f_i(\cdot)$ and $f_j(\cdot)$ exist and are strictly positive and continuous on $\eta_i \leq \eta_j$ where the η_i are unique medians, then η_{ij} is well defined for all N with $n_i > 0, n_j > 0$.

As a final preparation, we establish a result regarding the asymptotic dis-

tribution of the $(c - 1)$ -vector

$$\mathbf{S}_{iN} = n_i^{1/2} \begin{bmatrix} \hat{F}_i(\eta_{i1}) - F_i(\eta_{i1}) \\ \hat{F}_i(\eta_{i2}) - F_i(\eta_{i2}) \\ \vdots \\ \hat{F}_i(\eta_{i(i-1)}) - F_i(\eta_{i(i-1)}) \\ \hat{F}_i(\eta_{i(i+1)}) - F_i(\eta_{i(i+1)}) \\ \vdots \\ \hat{F}_i(\eta_{ic}) - F_i(\eta_{ic}) \end{bmatrix}.$$

There are c such independent vectors. The elements of \mathbf{S}_{iN} shall be indexed $1, \dots, i - 1, i + 1, \dots, c$, omitting the i th index.

Lemma 2.1 *Assume A2 and A3. As $N \rightarrow \infty$, \mathbf{S}_{iN} converges in law to a $(c - 1)$ -variate normal random variable with $\mathbf{0}$ mean vector and covariance matrix Δ_i ,*

$$(\Delta_i)_{jj'} = \begin{cases} F_i(\eta_{ij}^\circ) [1 - F_i(\eta_{ij}^\circ)] & j = j', \\ F_i(\min\{\eta_{ij}^\circ, \eta_{ij'}^\circ\}) - F_i(\eta_{ij}^\circ) F_i(\eta_{ij'}^\circ) & j \neq j'. \end{cases}$$

(The rows and columns of Δ_i again omit the index i .)

Proof: Without loss of generality assume $i = 1$, in which case \mathbf{S}_{1N} will have elements indexed by $j = 2, \dots, c$. Let $\boldsymbol{\alpha}$ be a $(c - 1)$ -vector (non-zero, elements $j = 2, \dots, c$), and define

$$Y_{Nk} = \sum_{j=2}^c \alpha_j I[X_{1k} \leq \eta_{1j}] \quad , \quad k = 1, \dots, n_1,$$

where $I(A)$ is the indicator function for the event A . For each $N \geq N_0$ form the row

of independent, identically distributed random variables Y_{Nk} , $k = 1, \dots, n_1$. Define

$$\mu_N = E[Y_{Nk}] = E \left[\sum_{j=2}^c \alpha_j I[X_{1k} \leq \eta_{1j}] \right] = \sum_{j=2}^c \alpha_j F_1(\eta_{1j}).$$

$$\begin{aligned} \sigma^2_N &= Var[Y_{Nk}] = Var \left[\sum_{j=2}^c \alpha_j I[X_{1k} \leq \eta_{1j}] \right] \\ &= \sum_{j=2}^c \alpha_j^2 F_1(\eta_{1j}) [1 - F_1(\eta_{1j})] \\ &\quad + 2 \sum_{j=2}^{c-1} \sum_{k=j+1}^c \alpha_j \alpha_k [F_1(\min\{\eta_{1j}, \eta_{1k}\}) - F_1(\eta_{1j}) F_1(\eta_{1k})]. \end{aligned}$$

Further, define

$$A_N = E \left[\sum_{k=1}^{n_1} Y_{Nk} \right] = n_1 \mu_N, \quad B_N^2 = Var \left[\sum_{k=1}^{n_1} Y_{Nk} \right] = n_1 \sigma^2_N.$$

By the Lindeberg central limit theorem for arrays of random random variables, independence within rows (Serfling (1980), pages 31–32), the asymptotic normality of $\sum_{k=1}^{n_1} Y_{Nk}$ can be established by showing, for arbitrary $\epsilon > 0$,

$$\lim_{N \rightarrow \infty} \frac{\sum_{k=1}^{n_1} \int_{|Y_{Nk} - \mu_N| > \epsilon B_N} (t - \mu_{Nk})^2 dG_{Nk}(t)}{\epsilon B_N^2} = 0,$$

where $G_{Nk}(t)$ is the distribution function for Y_{Nk} . For fixed N the Y_{Nk} are identically distributed Bernoulli trials, further, $Y_{N1}^2 I(|Y_{N1} - \mu_{N1}| > \epsilon B_N)$ is bounded above by $I(|Y_{N1} - \mu_{N1}| > \epsilon B_N)$. Hence

$$\begin{aligned} &\lim_{N \rightarrow \infty} \frac{\sum_{k=1}^{n_1} \int_{|Y_{Nk} - \mu_N| > \epsilon B_N} (t - \mu_{Nk})^2 dG_{Nk}(t)}{\epsilon B_N^2} \\ &= \lim_{N \rightarrow \infty} \frac{n_1 E \left[Y_{N1}^2 I(|Y_{N1} - \mu_N| > n_1^{1/2} \epsilon \sigma_N) \right]}{\epsilon n_1 \sigma_N^2} \end{aligned}$$

$$\begin{aligned}
&= \lim_{N \rightarrow \infty} \frac{E \left[Y_{N1}^2 I \left(|Y_{N1} - \mu_N| > n_1^{1/2} \epsilon \sigma_N \right) \right]}{\epsilon \sigma_N^2} \\
&\leq \lim_{N \rightarrow \infty} \frac{P \left[|Y_{N1} - \mu_N| > n_1^{1/2} \epsilon \sigma_N \right]}{\epsilon \sigma_N^2}.
\end{aligned} \tag{2.1}$$

Since $\lim_{N \rightarrow \infty} F_1(\eta_{ij}) = F_1(\eta_{ij}^\circ)$, it follows that

$$\begin{aligned}
\sigma_N^2 \rightarrow \sigma_1^2 &= \sum_{j=2}^c \alpha_j^2 F_1(\eta_{1j}^\circ) \left[1 - F_1(\eta_{1j}^\circ) \right] \\
&\quad + 2 \sum_{j=2}^{c-1} \sum_{k=j+1}^c \alpha_j \alpha_k \left[F_1(\min\{\eta_{1j}^\circ, \eta_{1k}^\circ\}) - F_1(\eta_{1j}^\circ) F_1(\eta_{1k}^\circ) \right] > 0.
\end{aligned}$$

Then $n_1^{1/2} \epsilon \sigma_N$ must diverge; for all N sufficiently large the probability in the numerator of (2.1) is 0, implying the limit is 0. The Lindeberg condition is established, therefore

$$\frac{\sum_{j=2}^c Y_{Nk} - A_N}{B_N} \xrightarrow{\mathcal{L}} Z \sim n(0, 1).$$

As $B_N/n_1^{1/2} \rightarrow \sigma_1$, it follows that

$$n_1^{1/2} \sum_{j=2}^c (Y_{Nk} - \mu_N) = \boldsymbol{\alpha}^T \mathbf{S}_{1N} \xrightarrow{\mathcal{L}} W \sim n(0, \sigma_1^2).$$

Since $\sigma_1^2 = \boldsymbol{\alpha}^T \Delta_1 \boldsymbol{\alpha}$, $\boldsymbol{\alpha}^T \mathbf{S}_{1N}$ has the limiting distribution of $\boldsymbol{\alpha}^T \mathbf{X}$, where \mathbf{X} is $(c-1)$ -variate normal, $\mathbf{0}$ mean and covariance matrix Δ_1 . Applying the Cramer-Wald device (Serfling (1980)), page 18), $\mathbf{S}_{1N} \xrightarrow{\mathcal{L}} \mathbf{X}$. ■

For each N the vectors \mathbf{S}_{iN} , $i = 1, \dots, c$, are independent, hence the limiting distribution of the collection of \mathbf{S}_{iN} , $i = 1, \dots, c$ follows immediately. Define the $c(c-1)$ -vector $\mathbf{S}_N = [\mathbf{S}_{1N}^T, \dots, \mathbf{S}_{cN}^T]^T$. The elements of \mathbf{S}_N will be doubly

indexed,

$$(\mathbf{S}_N)_{ij} = n_i^{1/2} [\hat{F}_i(\eta_{ij}) - F_i(\eta_{ij})], \quad (2.2)$$

where $1 \leq i \neq j \leq c$.

Corollary 2.1 *The limiting distribution of \mathbf{S}_N is $c(c-1)$ -variate normal with $\mathbf{0}$ mean and covariance matrix Δ ,*

$$\Delta = \begin{bmatrix} \Delta_1 & 0 & 0 & \cdots & \cdots & 0 \\ 0 & \Delta_2 & 0 & \cdots & \cdots & 0 \\ \vdots & & \ddots & \cdots & \cdots & \vdots \\ \vdots & & & & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \cdots & \Delta_c \end{bmatrix}.$$

2.2 An alternate expression for the Mood median statistic

We prove a generalization of the quantile representation theorem for non-identically distributed random variables. In what follows assume that conditions A2 and A3 hold. Without loss of generality the result is developed for $(i, j) = (1, 2)$.

Let $n = n_1 + n_2$, and define $\bar{F}_n(x) = [n_1 F_1(x) + n_2 F_2(x)]/n$. Note that for all $N > N_0$, $\bar{F}_n(\eta_{12}) = 1/2$.

Lemma 2.2 *For any real t , as $n \rightarrow \infty$ with $n_1/n \rightarrow \lambda_1/(\lambda_1 + \lambda_2)$,*

$$\frac{\bar{F}_n(\eta_{12} + t/n^{1/2}) - 1/2}{t/n^{1/2}} \rightarrow c_{12} > 0,$$

where

$$c_{12} = \frac{\lambda_1}{\lambda_1 + \lambda_2} f_1(\eta_{12}^\circ) + \frac{\lambda_2}{\lambda_1 + \lambda_2} f_2(\eta_{12}^\circ).$$

Proof: By definition

$$\begin{aligned} \frac{\bar{F}_n(\eta_{12} + t/n^{1/2}) - 1/2}{t/n^{1/2}} &= \\ &= \left(\frac{n_1}{n}\right) \frac{F_1(\eta_{12} + t/n^{1/2}) - F_1(\eta_{12})}{t/n^{1/2}} + \left(\frac{n_2}{n}\right) \frac{F_2(\eta_{12} + t/n^{1/2}) - F_2(\eta_{12})}{t/n^{1/2}}. \end{aligned}$$

By the continuity of $\eta_{12}(y)$ at $y = \lambda_1/(\lambda_1 + \lambda_2)$ and the differentiability of $F_1(x)$ in a neighborhood of η_{12}° , using Young's form of Taylor's theorem (Arnold (1990), page 232) results in

$$F_1(\eta_{12} + t/n^{1/2}) - F_1(\eta_{12}) = \frac{t f_1(\eta_{12})}{n^{1/2}} + o(|t/n^{1/2}|), \quad t/n^{1/2} \rightarrow 0. \quad (2.3)$$

Therefore,

$$\lim_{n \rightarrow \infty} \frac{F_1(\eta_{12} + t/n^{1/2}) - F_1(\eta_{12})}{t/n^{1/2}} = \lim_{n \rightarrow \infty} [f_1(\eta_{12}) + o(1)] = f_1(\eta_{12}^\circ), \quad (2.4)$$

as $f_1(\cdot)$ has been assumed continuous at η_{12}° . The remainder of the computation is entirely analogous; it follows that c_{12} is as given above. ■

The following technical device will be needed.

Lemma 2.3 (Ghosh (1971)) *Let U_n and W_n be two sequences of random variables satisfying the following conditions.*

C1. For all $\delta > 0$ there exists γ (depending on δ) such that $P(|W_n| > \gamma) < \delta$.

C2. For all t and all $\epsilon > 0$

$$\lim_{n \rightarrow \infty} P(U_n \leq t, W_n \geq t + \epsilon) = 0, \quad (2.5)$$

$$\lim_{n \rightarrow \infty} P(U_n \geq t + \epsilon, W_n \leq t) = 0. \quad (2.6)$$

Then $U_n - W_n \xrightarrow{p} 0$ as $n \rightarrow \infty$.

The following theorem is a version of the quantile representation theorem for the case of two non-identically distributed samples. The proof is based on methods used by Ghosh (1971) in deriving a representation theorem (*i.i.d.* case) under more relaxed regularity conditions than those stated in the result of Bahadur (1966).

Theorem 2.1 As $n \rightarrow \infty$ with $n_1/n \rightarrow \lambda_1/(\lambda_1 + \lambda_2)$,

$$(\hat{\eta}_{12} - \eta_{12}) = -\frac{\hat{F}_{12}(\eta_{12}) - 1/2}{c_{12}} + o_p(n^{-1/2}). \quad (2.7)$$

Proof: Recall that $\hat{\eta}_{12}$ has been defined to be the smallest sample median, i.e. $\hat{\eta}_{12} = \inf\{x : \hat{F}_{12}(x) \geq 1/2\}$. Then for any real number t ,

$$\begin{aligned} \{\omega : n^{1/2}(\hat{\eta}_{12} - \eta_{12}) \leq t\} &= \{\omega : 1/2 \leq F_n(\eta_{12} + t/n^{1/2})\} \\ &= \left\{ \omega : n^{1/2} \frac{[\bar{F}_n(\eta_{12} + t/n^{1/2}) - F_n(\eta_{12} + t/n^{1/2})]}{c_{12}} \leq t_n \right\}, \end{aligned}$$

where

$$t_n = \frac{n^{1/2}}{c_{12}} [\bar{F}_n(\eta_{12} + t/n^{1/2}) - 1/2].$$

Note that $\lim_{n \rightarrow \infty} t_n = t$ (Lemma 2.2). Set

$$Q_{t,n} = n^{1/2} \frac{[\bar{F}_n(\eta_{12} + t/n^{1/2}) - F_n(\eta_{12} + t/n^{1/2})]}{c_{12}},$$

$$W_n = n^{1/2} \frac{[\bar{F}_n(\eta_{12}) - F_n(\eta_{12})]}{c_{12}}.$$

Then

$$W_n - Q_{t,n} = \frac{n^{1/2}}{c_{12}} \left\{ [F_n(\eta_{12} + t/n^{1/2}) - F_n(\eta_{12})] - [\bar{F}_n(\eta_{12} + t/n^{1/2}) - \bar{F}_n(\eta_{12})] \right\}.$$

With $N > N_0$, $E[W_n - Q_{t,n}] = 0$. Further, $E[(W_n - Q_{t,n})^2] = q_n(1 - q_n)/c_{12}^2$, where $q_n = |\bar{F}_n(\eta_{12} + t/n^{1/2}) - \bar{F}_n(\eta_{12})|$. From Lemma 2.2, $q_n \rightarrow 0$, implying that $\lim_{n \rightarrow \infty} E[(W_n - Q_{t,n})^2] = 0$, hence $(W_n - Q_{t,n}) \xrightarrow{p} 0$. By definition W_n can be expressed

$$W_n = -\left(\frac{n}{n_1}\right) n_1^{1/2} [\hat{F}_1(\eta_{12}) - F_1(\eta_{12})] - \left(\frac{n}{n_2}\right) n_2^{1/2} [\hat{F}_2(\eta_{12}) - F_2(\eta_{12})].$$

The asymptotic normality of W_n is established by Lemma 2.1 and the convergence of the sample proportions. W_n is bounded in probability; C1 of Lemma 2.3 has been established.

Take $U_n = n^{1/2}(\hat{\eta}_{12} - \eta_{12})$. To establish condition C2 let t and $\epsilon > 0$ be arbitrary. Then

$$\begin{aligned} P[U_n \leq t, W_n \geq t + \epsilon] &= P[Q_{t,n} \leq t_n, W_n \geq t + \epsilon] \\ &\leq P[(W_n - Q_{t,n}) \geq (t - t_n) + \epsilon] \end{aligned}$$

$$\leq P[|W_n - Q_{t,n}| \geq (t - t_n) + \epsilon]. \quad (2.8)$$

As $n \rightarrow \infty$, $t_n \rightarrow t$ and $(W_n - Q_{t,n}) \xrightarrow{p} 0$, hence the probability in (2.8) converges to 0. Hence (2.5) is satisfied. A similar argument establishes (2.6), therefore C2 has been verified.

Having established both C1 and C2, Lemma 2.3 can be applied to state that $U_n - W_n \xrightarrow{p} 0$. This is equivalent to the statement of the theorem. ■

Corollary 2.2 *Assuming A2 and A3, the limiting distribution of $n^{1/2}(\hat{\eta}_{12} - \eta_{12})$ is normal with mean 0 and variance*

$$\frac{1}{c_{12}^2} \left\{ \frac{\lambda_1}{\lambda_1 + \lambda_2} F_1(\eta_{12}^\circ) [1 - F_1(\eta_{12}^\circ)] + \frac{\lambda_2}{\lambda_1 + \lambda_2} F_2(\eta_{12}^\circ) [1 - F_2(\eta_{12}^\circ)] \right\}.$$

Proof: From Theorem 2.1,

$$\begin{aligned} n^{1/2}(\hat{\eta}_{12} - \eta_{12}) &= -n^{1/2} \left[\frac{\hat{F}_{12}(\eta_{12}) - 1/2}{c_{12}} \right] + o_p(1) \\ &= - \left(\frac{n_1}{n} \right)^{1/2} c_{12}^{-1} n_1^{1/2} [\hat{F}_1(\eta_{12}) - F_1(\eta_{12})] \\ &\quad - \left(\frac{n_2}{n} \right)^{1/2} c_{12}^{-1} n_2^{1/2} [\hat{F}_2(\eta_{12}) - F_2(\eta_{12})] + o_p(1) \end{aligned}$$

Convergence of the sample proportions, the result of Lemma 2.1 and Slutsky's theorem proves the corollary. ■

The following result is given in Hettmansperger (1984), pages 78–79.

Lemma 2.4 *Suppose that $W_n(b) = U_n(b) + c_n b$ where $U_n(b)$ is monotone in b and $|c_n| \leq c < \infty$. Suppose that for each b , $W_n(b) \xrightarrow{p} 0$ as $n \rightarrow \infty$. Then for any $B > 0$*

and any $\epsilon > 0$,

$$\lim_{n \rightarrow \infty} P \left\{ \sup_{|b| \leq B} |W_n(b)| > \epsilon \right\} = 0.$$

Lemma 2.5 *Assume A2 and A3. For any $\epsilon > 0$ and any $B > 0$, as $n \rightarrow \infty$,*

$$P \left\{ \sup_{|b| \leq B} \left| n^{1/2} \left[\hat{F}_1(\eta_{12} + b/n^{1/2}) - \hat{F}_1(\eta_{12}) \right] - bf_1(\eta_{12}^\circ) \right| > \epsilon \right\} \rightarrow 0.$$

Proof: Set

$$U_n(b) = n^{1/2} \left[\hat{F}_1(\eta_{12} + b/n^{1/2}) - \hat{F}_1(\eta_{12}) \right], \quad c_n = -f_1(\eta_{12}^\circ).$$

Define $W_n = U_n(b) + c_n b$; U_n is non-decreasing in b and c_n is bounded. From (2.4), as $n \rightarrow \infty$,

$$E[U_n(b)] = n^{1/2} \left[F_1(\eta_{12} + b/n^{1/2}) - F_1(\eta_{12}) \right] \rightarrow bf_1(\eta_{12}^\circ),$$

$$\text{Var}[U_n(b)] = \left\{ F_1(\eta_{12} + b/n^{1/2}) - F_1(\eta_{12}) \right\} \times$$

$$\left\{ 1 - \left[F_1(\eta_{12} + b/n^{1/2}) - F_1(\eta_{12}) \right] \right\} \rightarrow 0.$$

This verifies that $U_n(b) \xrightarrow{p} bc_n$. Applying Lemma 2.4 establishes the lemma. ■

Lemma 2.6 *Assume A2 and A3. As $n \rightarrow \infty$,*

$$n^{1/2} \left[\hat{F}_1(\hat{\eta}_{12}) - F_1(\eta_{12}) \right] =$$

$$n^{1/2} \left[\hat{F}_1(\eta_{12}) - F_1(\eta_{12}) \right] + f_1(\eta_{12}^\circ) n^{1/2} (\hat{\eta}_{12} - \eta_{12}) + o_p(1). \quad (2.9)$$

Proof: Let $\epsilon > 0$ and $\delta > 0$. From Corollary 2.2 we have $n^{1/2} (\hat{\eta}_{12} - \eta_{12})$ bounded in probability, there then exist B and $N_{1\delta}$ such that $P[|n^{1/2} (\hat{\eta}_{12} - \eta_{12})| > B] \leq \delta/2$

for all $N > N_{1\delta}$. Further, for B as given, there exists $N_{2\delta}$ such that for all $N > N_{2\delta}$

$$P \left\{ \sup_{|b| \leq B} \left| n^{1/2} \left[\hat{F}_1(\eta_{12} + b/n^{1/2}) - \hat{F}_1(\eta_{12}) \right] - bf(\eta_{12}^\circ) \right| > \epsilon \right\} < \frac{\delta}{2}.$$

For $N > \max\{N_{1\delta}, N_{2\delta}\}$,

$$\begin{aligned} & P \left\{ \left| n^{1/2} \left[\hat{F}_1(\hat{\eta}_{12}) - \hat{F}_1(\eta_{12}) \right] - f(\eta_{12}^\circ) n^{1/2} (\hat{\eta}_{12} - \eta_{12}) \right| > \epsilon \right\} \\ & \leq P \left\{ \sup_{|b| \leq B} \left| n^{1/2} \left[\hat{F}_1(\eta_{12} + b/n^{1/2}) - \hat{F}_1(\eta_{12}) \right] - f(\eta_{12}^\circ) b \right| > \epsilon \right\} \\ & \quad + P \left\{ \left| n^{1/2} (\hat{\eta}_{12} - \eta_{12}) \right| > B \right\} \\ & \leq \frac{\delta}{2} + \frac{\delta}{2} = \delta. \end{aligned}$$

Hence $n^{1/2} [\hat{F}_1(\hat{\eta}_{12}) - \hat{F}_1(\eta_{12})] - f(\eta_{12}^\circ) [n^{1/2} (\hat{\eta}_{12} - \eta_{12})] \xrightarrow{p} 0$. ■

We now give a re-expression for the Mood statistic under the alternate hypothesis. For each pair $(i, j) : 1 \leq i < j \leq c$ define

$$v_{ijN} = (n_i + n_j)n_i \hat{F}_i(\hat{\eta}_{ij}) - \mu_{ijN}, \text{ where } \mu_{ijN} = (n_i + n_j)n_i F_i(\eta_{ij}).$$

Let \mathbf{V}_N be the d -vector of v_{ijN} .

Theorem 2.2 *Assuming A2 and A3, for each $(i, j) : 1 \leq i < j \leq c$, as $n_i + n_j \rightarrow \infty$,*

$$\begin{aligned} N^{-3/2} v_{ijN} &= \left[\frac{n_i + n_j}{N} - \frac{n_i f_i(\eta_{ij}^\circ)}{N c_{ij}} \right] \left(\frac{n_i}{N} \right)^{1/2} n_i^{1/2} \left[\hat{F}_i(\eta_{ij}) - F_i(\eta_{ij}) \right] \\ &\quad - \left[\frac{n_i f_i(\eta_{ij}^\circ)}{N c_{ij}} \right] \left(\frac{n_j}{N} \right)^{1/2} n_j^{1/2} \left[\hat{F}_j(\eta_{ij}) - F_j(\eta_{ij}) \right] + o_p(1). \end{aligned} \quad (2.10)$$

Proof: Substituting (2.7) into (2.9) yields the result. ■

Under H_0 it is sufficient to take $F(\cdot)$ differentiable at η ; the result is then

equivalent to that given in Theorem 1.1; convergence with probability one has been replaced with convergence in probability.

2.3 Multivariate limiting normality

From (2.10) it follows that

$$N^{-3/2}\mathbf{V}_N = A_N B_N \mathbf{S}_N + \mathbf{R}_N,$$

where A_N , B_N , \mathbf{S}_N and \mathbf{R}_N are as follows.

- A_N is a $d \times 2d$ matrix with components

$$(A_N)_{(ij)(ij)} = \left[\frac{n_i + n_j}{N} - \frac{n_i f_i(\eta_{ij}^\circ)}{N c_{ij}} \right], \quad (2.11)$$

$$(A_N)_{(ij)(ji)} = \left[-\frac{n_i f_i(\eta_{ij}^\circ)}{N c_{ij}} \right]. \quad (2.12)$$

All other terms of A_N are 0. The index set is $1 \leq i < j \leq c$, $1 \leq k \neq l \leq c$. The structure of A_N for the case $c = 5$ is displayed in Figure 2.1.

- B_N is a $2d \times 2d$ diagonal matrix, $(B_N)_{(ij)(ij)} = (n_i/N)^{1/2}$. The rows and columns of B_N are ordered exactly as are the columns of A_N .
- \mathbf{S}_N is as defined in (2.2).
- \mathbf{R}_N is a $2d$ -vector of terms $o_p(1)$.

Theorem 2.3 *Under A2 and A3, as $N \rightarrow \infty$ the vector $N^{-3/2}\mathbf{V}_N$ converges in law to a d -variate normal random vector with $\mathbf{0}$ mean and covariance matrix Σ ,*

	<u>Column</u>																			
<u>Row</u>	12	13	14	15	21	23	24	25	31	32	34	35	41	42	43	45	51	52	53	54
12	+	0	0	0	-	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
13	0	+	0	0	0	0	0	0	-	0	0	0	0	0	0	0	0	0	0	0
14	0	0	+	0	0	0	0	0	0	0	0	0	-	0	0	0	0	0	0	0
15	0	0	0	+	0	0	0	0	0	0	0	0	0	0	0	0	-	0	0	0
23	0	0	0	0	0	+	0	0	0	-	0	0	0	0	0	0	0	0	0	0
24	0	0	0	0	0	0	+	0	0	0	0	0	0	-	0	0	0	0	0	0
25	0	0	0	0	0	0	0	+	0	0	0	0	0	0	0	0	0	-	0	0
34	0	0	0	0	0	0	0	0	0	0	+	0	0	0	-	0	0	0	0	0
35	0	0	0	0	0	0	0	0	0	0	0	+	0	0	0	0	0	0	-	0
45	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	+	0	0	0	-

A “+” in the figure denotes an entry with row and column labels the same: see (2.11); also (2.13) and (3.3).

A “-” in the figure denotes an entry with row and column labels the reverse of each other: see (2.12); also (2.14) and (3.4).

Figure 2.1: Structure of the A Matrix Under H_1

elements $\sigma_{(ij)(km)}$,

$$\begin{aligned} \sigma_{(ij)(ij)} &= \lambda_i \lambda_j \left\{ \lambda_j f_j^2(\eta_{ij}^\circ) F_i(\eta_{ij}^\circ) [1 - F_i(\eta_{ij}^\circ)] \right. \\ &\quad \left. + \lambda_i f_i^2(\eta_{ij}^\circ) F_j(\eta_{ij}^\circ) [1 - F_j(\eta_{ij}^\circ)] \right\} / c_{ij}^2, \\ \sigma_{(ij)(im)} &= \frac{\lambda_i \lambda_j \lambda_m f_j(\eta_{ij}^\circ) f_m(\eta_{im}^\circ) [F_i(\min\{\eta_{ij}^\circ, \eta_{im}^\circ\}) - F_i(\eta_{ij}^\circ) F_i(\eta_{im}^\circ)]}{c_{ij} c_{im}}, \\ \sigma_{(ij)(kj)} &= \frac{\lambda_i \lambda_j \lambda_k f_i(\eta_{ij}^\circ) f_k(\eta_{kj}^\circ) [F_j(\min\{\eta_{ij}^\circ, \eta_{kj}^\circ\}) - F_j(\eta_{ij}^\circ) F_j(\eta_{kj}^\circ)]}{c_{ij} c_{kj}}, \\ \sigma_{(ij)(jm)} &= \frac{-\lambda_i \lambda_j \lambda_m f_i(\eta_{ij}^\circ) f_m(\eta_{jm}^\circ) [F_j(\min\{\eta_{ij}^\circ, \eta_{jm}^\circ\}) - F_j(\eta_{ij}^\circ) F_j(\eta_{jm}^\circ)]}{c_{ij} c_{jm}}, \\ \sigma_{(ij)(km)} &= 0, \end{aligned}$$

where the indices i, j, k, m are distinct, $1 \leq i < j \leq c$, $1 \leq l < m \leq c$.

Proof: Convergence of sample proportions implies that A_N converges component-wise to the matrix A given below:

$$(A)_{(ij)(ij)} = \lim_{N \rightarrow \infty} (A_N)_{(ij)(ij)} = \lim_{N \rightarrow \infty} \left[\frac{n_i + n_j}{N} - \frac{n_i f_i(\eta_{ij}^\circ)}{N c_{ij}} \right] = \frac{\lambda_j f_j(\eta_{ij}^\circ)}{c_{ij}}, \quad (2.13)$$

$$(A)_{(ij)(ji)} = \lim_{N \rightarrow \infty} (A_N)_{(ij)(ji)} = \lim_{N \rightarrow \infty} - \left[\frac{n_i f_i(\eta_{ij}^\circ)}{N c_{ij}} \right] = - \frac{\lambda_i f_i(\eta_{ij}^\circ)}{c_{ij}}, \quad (2.14)$$

where the remaining elements of A are 0. (The structure of A is displayed in Figure 2.1, page 30.) $B_N \rightarrow B$, diagonal with elements $(B)_{(ij)(ij)} = \lambda_i^{1/2}$. From Lemma 2.1, \mathbf{S}_N is limiting normal with $\mathbf{0}$ mean and covariance matrix Δ . By Slutsky's theorem:

$$N^{-3/2} \mathbf{V}_N = A_N B_N \mathbf{S}_N + \mathbf{R}_N \xrightarrow{\mathcal{L}} \mathbf{X},$$

where \mathbf{X} is a d -variate normal random variable having mean vector $AB\mathbf{0} = \mathbf{0}$, and covariance matrix given by $\Sigma = AB\Delta B^T A^T$. Multiplying the matrices yields the solution stated above. ■

Finally, let \mathbf{b} be a d -vector of weights. An immediate consequence of Theorem 2.3 is that any linear combination of the $N^{-3/2}v_{ijN}$ is normal in the limit. As before let $V_N(\mathbf{b}) = \mathbf{b}^T \mathbf{V}_N = N^{-3/2} \sum_{i=1}^{c-1} \sum_{j=i+1}^c b_{ij} v_{ijN}$.

Corollary 2.3 *Under the conditions given in Theorem 2.3, $V_N(\mathbf{b})$ is limiting normal with mean 0 and variance $\mathbf{b}^T \Sigma \mathbf{b}$.*

Chapter 3

Asymptotic Relative Efficiency

In this chapter we develop the asymptotic relative efficiency (Pitman, 1949) of the test based on V_N with respect to competing statistics. The idea of asymptotic relative efficiency is to choose a sequence of alternative hypotheses H_1^M which vary with the sample sizes in such a manner that the powers of two tests for this sequence of alternatives have a common limit less than 1. The comparison of the two tests is then made on a sample size basis.

Suppose that two tests V_N and $V_{N'}$ require N and N' observations, respectively, to attain the power β at level of significance α for the hypothesis H_0 against hypothesis H_1^M . The asymptotic relative efficiency of V_N with respect to $V_{N'}$ is defined to be

$$e_{V, V'} = \lim_{M \rightarrow \infty} \frac{N'}{N}.$$

Let \mathcal{L} denote the class of linear combinations of Mood statistics $V_N(\mathbf{b}) = N^{-3/2} \mathbf{b}^T \mathbf{V}_N = N^{-3/2} \sum_{i < j} b_{ij} (n_i + n_j) [M_{ij} - n_i/2]$, where the b_{ij} are weighting coefficients. Similarly take \mathcal{L}^* to be the subset of \mathcal{L} consisting of linear combinations of Mood statistics on adjacent samples only, $V_N(\mathbf{b}^*) = \sum_{i=1}^{c-1} b_i^* (n_i + n_{i+1}) [M_{i(i+1)} - n_i/2]$. We show that for each statistic $V_N(\mathbf{b})$ in \mathcal{L} (i.e., each vector of weighting coefficients \mathbf{b}) there exists a vector of coefficients \mathbf{b}^* and statistic $V_N(\mathbf{b}^*)$ in \mathcal{L}^* such that the two tests differ under H_0 by terms negligible in probability and have an asymptotic relative efficiency of unity under the appropriate sequence of alternatives H_1^M . The construction of \mathbf{b}^* is provided. For linear combinations of Chernoff and Savage statis-

tics, and assuming equal sample sizes within groups, Tryon and Hettmansperger (1973) exhibit the same result.

The tests in \mathcal{L}^* have $(c-1)$ “free” coefficients, b_1, \dots, b_{c-1} . In the c -sample translation model there are $(c-1)$ shifts in the population locations. A natural question is whether the coefficients can be selected in such a way as to optimize the test for a particular set of location shifts. Using asymptotic relative efficiency as the criteria, optimally weighted tests are constructed within the class \mathcal{L}^* .

3.1 Asymptotic distribution under a sequence of shift alternatives

Restrict Ω , the class of admissible hypotheses, to the class of translation alternatives: $F_i(x) = F(x - \theta_i)$, for $i = 1, \dots, c$, where $\theta_1 \leq \dots \leq \theta_c$, with, for some i , $\theta_i < \theta_{i+1}$, for some arbitrary choice of F , where $F(x)$ has a strictly positive, continuous density $f(x)$ in some neighborhood of the point $\eta : F(\eta) = 1/2$. We develop the distribution of V_N assuming a sequence of alternative hypotheses H_1^M for $M = 1, 2, \dots$. The hypothesis H_1^M specifies, for each $i = 1, \dots, c$, that $F_i(x) = F(x - \theta_{iM})$ where the θ_i are nondecreasing and not all the same, and $\theta_{iM} = \theta_i/\sqrt{M}$. The letter M is used to index a sequence of situations in which H_1^M is the true hypothesis. A limiting distribution is found as $M \rightarrow \infty$ assuming that N is a function of M , $N = N(M)$, with $N/M \rightarrow \rho_V < \infty$.

Define $\mu_{ijN}(\boldsymbol{\theta}) = (n_i + n_j)n_i F_i(\eta_{ij})$, where η_{ij} is computed under the assumption that the hypothesis H_1^M is true ($\mu_{ijN}(\mathbf{0}) = (n_i + n_j)n_i/2$, the null value). Let $\boldsymbol{\mu}_N(\boldsymbol{\theta})$ be the d -vector of $\mu_{ijN}(\boldsymbol{\theta})$.

Theorem 3.1 *For each index M assume that the hypothesis H_1^M is true. Suppose $F(\cdot)$ has strictly positive, continuous density in some neighborhood of η . Then the*

limiting distribution of the vector $N^{-3/2} \mathbf{V}_N$ is d -variate normal with mean $\mathbf{0}$ and covariance matrix Σ as specified by the null distribution (1.14):

$$\sigma_{(ij)(km)} = \begin{cases} \frac{\lambda_i \lambda_j (\lambda_i + \lambda_j)}{4} & i = k, j = m, \\ \frac{\lambda_i \lambda_j \lambda_m}{4} & i = k, j \neq m, \\ \frac{\lambda_i \lambda_j \lambda_k}{4} & i \neq k, j = m, \\ -\frac{\lambda_i \lambda_j \lambda_m}{4} & j = k, \\ 0 & \text{otherwise.} \end{cases}$$

Proof: The proof is a specification of Theorem 2.2 , making use of the fact that both η_{ij} and η_{ij}° converge to η as $M \rightarrow \infty$. Hence $F(\eta_{ij}) \rightarrow 1/2$ and $f(\eta_{ij}^\circ) \rightarrow f(\eta)$. This is proved with a stronger result to be implemented later.

Let δ be arbitrarily small and assume that $F_i(x) = F(x - \theta_i \delta)$. Then η_{ij} satisfies

$$H(\eta_{ij}, \delta) = \frac{n_i}{n_i + n_j} F(\eta_{ij} - \theta_i \delta) + \frac{n_j}{n_i + n_j} F(\eta_{ij} - \theta_j \delta) - \frac{1}{2} = 0. \quad (3.1)$$

Since H has continuous partial derivatives with respect to both η_{ij} and δ in some neighborhood of $\eta_{ij} = \eta, \delta = 0$, and

$$\left. \frac{\partial}{\partial \eta_{ij}} H(\eta_{ij}, \delta) \right|_{(\eta, 0)} = f(\eta) > 0,$$

the implicit function theorem can again be applied. There exists a function $\eta(\delta)$ having continuous first derivative such that $\eta_{ij} = \eta(\delta)$ satisfies $H(\eta_{ij}, \delta) = 0$ for δ in some neighborhood of 0 with $\eta(0) = \eta$. Implicit differentiation of the relation (3.1)

gives

$$\left. \frac{d}{d\delta} \eta(\delta) \right|_{\delta=0} = \frac{n_i}{n_i + n_j} \theta_i + \frac{n_j}{n_i + n_j} \theta_j.$$

Using Young's form of Taylor's theorem,

$$F_i(\eta_{ij}) = F(\eta(\delta) - \theta_i \delta) = \frac{1}{2} + f(\eta) \left[\frac{n_j}{n_i + n_j} \right] (\theta_j - \theta_i) \delta + o(|\delta|), \delta \rightarrow 0. \quad (3.2)$$

Take $\delta = M^{-1/2}$, then as $M \rightarrow \infty$, $F_i(\eta_{ij}) \rightarrow 1/2 = F(\eta)$. Applying this in the proof of Lemma 2.1 results in the vector \mathbf{S}_N having limiting $c(c-1)$ -variate normal distribution with mean $\mathbf{0}$ and covariance matrix Δ of the form specified in Corollary 2.1, where the non-zero components of Δ are each equal to $1/4$.

In a similar fashion the continuous function $\eta^\circ(\delta)$ solves

$$\frac{\lambda_i}{\lambda_i + \lambda_j} F(\eta^\circ(\delta) - \theta_i \delta) + \frac{\lambda_j}{\lambda_i + \lambda_j} F(\eta^\circ(\delta) - \theta_j \delta) = \frac{1}{2}.$$

With $\delta = M^{-1/2}$, this continuity implies that as $M \rightarrow \infty$, $\eta_{ij}^\circ = \eta^\circ(\delta) \rightarrow \eta$. The continuity of $f(x)$ at η gives $\lim_{M \rightarrow \infty} f(\eta_{ij}^\circ) = f(\eta)$. In the proof of Theorem 2.2 we have

$$N^{-3/2} \mathbf{V}_N = A_N B_N \mathbf{S}_N + \mathbf{R}_N.$$

As $M \rightarrow \infty$ the matrix A_N converges to A ,

$$(A)_{(ij)(ij)} = \lambda_j, \quad (3.3)$$

$$(A)_{(ij)(ji)} = -\lambda_i, \quad (3.4)$$

and terms 0 elsewhere. (The structure of A is displayed in Figure 2.1, page 30.)

By Slutsky's theorem the distribution of $N^{-3/2}\mathbf{V}_N$ is then d -variate normal with mean $\mathbf{0}$ and covariance matrix $AB\Delta B^T A^T = \Sigma$ ■

3.2 Asymptotic relative efficiency

We first compute the asymptotic relative efficiency of the test based on V_N relative to the normal theory competitor.

Taking $\delta = M^{-1/2}$, by (3.2),

$$\begin{aligned}
& \lim_{M \rightarrow \infty} N^{-3/2} [\mu_{ijN}(\boldsymbol{\theta}) - \mu_{ijN}(\mathbf{0})] \\
&= \lim_{M \rightarrow \infty} N^{-3/2} (n_i + n_j) n_i \left[F_i(\eta_{ij}) - \frac{1}{2} \right] \\
&= \lim_{M \rightarrow \infty} N^{-3/2} (n_i + n_j) n_i \left[\left(\frac{n_j}{n_i + n_j} \right) f(\eta) (\theta_j - \theta_i) \frac{1}{\sqrt{M}} + o(|M^{-1/2}|) \right] \\
&= \lim_{M \rightarrow \infty} \left[\frac{n_i n_j}{N^2} \left(\frac{N}{M} \right)^{1/2} f(\eta) (\theta_j - \theta_i) + o(1) \right] \\
&= (\rho_V)^{1/2} \lambda_i \lambda_j f(\eta) (\theta_j - \theta_i). \tag{3.5}
\end{aligned}$$

Assuming the sequence of alternatives H_1^M , the limiting distribution of the test statistic can now be established.

Theorem 3.2 *For each index M assume that the hypothesis H_1^M is true. Suppose $F(x)$ has strictly positive, continuous density $f(x)$ in some neighborhood of η . As $M \rightarrow \infty, N/M \rightarrow \rho_V$, the limiting distribution of*

$$N^{-3/2} \sum_{i=1}^{c-1} \sum_{j=i+1}^c b_{ij} (n_i + n_j) (M_{ij} - n_i/2)$$

$$= N^{-3/2} \sum_{i=1}^{c-1} \sum_{j=i+1}^c b_{ij} [(n_i + n_j) n_i \hat{F}_i(\hat{\eta}_{ij}) - \mu_{ijN}(0)]$$

is normal with mean

$$(\rho_V)^{1/2} f(\eta) \sum_{i=1}^{c-1} \sum_{j=i+1}^c b_{ij} \lambda_i \lambda_j (\theta_j - \theta_i) \quad (3.6)$$

and variance $\mathbf{b}^T \Sigma \mathbf{b}$.

Proof: Write

$$\begin{aligned} N^{-3/2} \sum_{i=1}^{c-1} \sum_{j=i+1}^c b_{ij} [(n_i + n_j) n_i \hat{F}_i(\hat{\eta}_{ij}) - \mu_{ijN}(0)] = \\ N^{-3/2} \mathbf{b}^T \mathbf{V}_N + N^{-3/2} \sum_{i=1}^{c-1} \sum_{j=i+1}^c b_{ij} [\mu_{ijN}(\boldsymbol{\theta}) - \mu_{ijN}(\mathbf{0})] + o_p(1). \end{aligned} \quad (3.7)$$

By Theorem 3.1, the first term of (3.7) is limiting normal with mean 0 and variance $\mathbf{b}^T \Sigma \mathbf{b}$. Applying (3.5) and Slutsky's theorem proves the result. ■

The normal theory competitor is based on the statistic

$$T_N = N^{-3/2} \sum_{i=1}^{c-1} \sum_{j=i+1}^c b_{ij} t_{ijN},$$

where $t_{ijN} = n_i n_j (\bar{X}_j - \bar{X}_i)$, with $\bar{X}_i = \sum_{l=1}^{n_i} X_{il} / n_i$. Assume that $\sigma^2 = \text{Var}(X_{il})$ is finite. For the test based on T_N let $N = N(M)$ with $N/M \rightarrow \rho_T < \infty$. Under the sequence of translation alternatives specified above with $F_i(x) = F(x - \theta_{iM})$ it can be shown (see Puri (1965)) that the limiting distribution of T_N is normal with mean

$$(\rho_T)^{1/2} \sum_{i=1}^{c-1} \sum_{j=i+1}^c b_{ij} \lambda_i \lambda_j (\theta_j - \theta_i), \quad (3.8)$$

and variance $\mathbf{b}^T \Gamma \mathbf{b}$, where Γ is the $d \times d$ matrix

$$(\Gamma)_{(ij)(km)} = \begin{cases} \sigma^2 \lambda_i \lambda_j (\lambda_i + \lambda_j) & i = k, j = m, \\ \sigma^2 \lambda_i \lambda_j \lambda_m & i = k, j \neq m, \\ \sigma^2 \lambda_i \lambda_j \lambda_k & i \neq k, j = m, \\ \sigma^2 \lambda_i \lambda_j \lambda_m & j = k \\ 0 & \text{otherwise.} \end{cases}$$

Corollary 3.1 *Suppose $F(\cdot)$ is such that $F(\eta) = 1/2$ is unique, $F(\cdot)$ has strictly positive, continuous density $f(\cdot)$ in some neighborhood of η , and that $F(\cdot)$ has finite first moment ξ and finite variance σ^2 . Then the asymptotic relative efficiency of the test based on V relative to the test based on T is*

$$e_{V,T}(F) = 4\sigma^2 f^2(\eta). \quad (3.9)$$

Proof: Let N' be the sample size for the test based on T , $N'/M \rightarrow \rho_T$. In order for the two tests to achieve the same power under the sequence H_1^M , the non-centrality parameters (3.6) and (3.8) are equated:

$$\frac{(\rho_V)^{1/2} f(\eta) \sum_{i=1}^{c-1} \sum_{j=i+1}^c b_{ij} \lambda_i \lambda_j (\theta_j - \theta_i)}{[\mathbf{b}^T \Sigma \mathbf{b}]^{1/2}} = \frac{(\rho_T)^{1/2} \sum_{i=1}^{c-1} \sum_{j=i+1}^c b_{ij} \lambda_i \lambda_j (\theta_j - \theta_i)}{[\mathbf{b}^T \Gamma \mathbf{b}]^{1/2}}. \quad (3.10)$$

To complete the evaluation, write

$$e_{V,T}(F) = \lim_{M \rightarrow \infty} \frac{N'}{N} = \lim_{M \rightarrow \infty} \frac{N'(M)}{N(M)} = \lim_{M \rightarrow \infty} \frac{\left(\frac{N'(M)}{M}\right)}{\left(\frac{N(M)}{M}\right)} = \frac{\rho_T}{\rho_V}.$$

This ratio is found by solving (3.10) for ρ_T/ρ_V (using $\Gamma = 4\sigma^2\Sigma$), and evaluates to (3.9). ■

Now consider a pairwise comparison statistic based on Chernoff-Savage

statistics. Let J_n be a (score) function, defined at $1/n, 2/n, \dots, n/n$, and extended to $(0, 1]$ by letting it be constant on $[\nu/n, (\nu + 1)/n]$. Letting

$$w_{ij} = \int_{-\infty}^{+\infty} J_{n_i+n_j} [\hat{F}_{ij}(x)] d(\hat{F}_i(x) - \hat{F}_j(x)),$$

then the statistic

$$W_N = N^{-3/2} \sum_{i=1}^{c-1} \sum_{j=i+1}^c b_{ij} n_i n_j w_{ij}$$

is a competitor to V_N . Typically the function $J_n[\nu/(n)]$ is taken to be the expected value of the ν -th order statistic of a sample of size n from some continuous distribution function or the ν/n -th quantile of that distribution function. For the Mann-Whitney rank statistic, $J(u) = u$, $0 < u < 1$. The normal-scores statistic is generated by $J(u) = \Phi^{-1}(u)$, $0 < u < 1$. Puri (1965) makes the following assumptions.

B1. $\lim_{N \rightarrow \infty} J_N(u) = J(u)$ exists for $0 < u < 1$ and is not a constant.

B2. $\int_{I_{n_i+n_j}} \{J_{n_i+n_j} [\hat{F}_{ij}(x)] - J[\hat{F}_{ij}(x)]\} d\hat{F}_i(x) = o_p(N^{-1/2})$, where $I_{n_i+n_j} = \{x : 0 < \hat{F}_{ij}(x) < 1\}$, $1 \leq i < j \leq c$.

B3. $J_N(1) = o(N^{1/2})$.

B4. $|J^{(i)}(u)| = |d^{(i)}J/du^{(i)}| \leq K[u(1-u)]^{\delta-1/2-i}$, $i = 0, 1, 2$, for some K and some $\delta > 0$.

In addition, assume the following hold (Hodges & Lehmann, 1956).

B5. $F(\cdot)$ is a continuous distribution function, differentiable in each of the open intervals $(-\infty, a_1), (a_1, a_2), \dots, (a_{k-1}, a_k), (a_k, \infty)$ and the derivative of $F(\cdot)$ is bounded in each of these intervals.

B6. The function $(d/dx) J[F(x)]$ is bounded as $x \rightarrow \pm\infty$.

Lemma 3.1 (Puri (1965)) *Assume B1 - B6 hold. For each index M assume that $N = N(M)$ such that $N/M \rightarrow \rho_W > 0$, and that the hypothesis H_1^M is true. Then the statistic W_N has a limiting distribution with mean*

$$\rho_W^{1/2} \sum_{i=1}^{c-1} \sum_{j=i+1}^c b_{ij} \lambda_i \lambda_j (\theta_j - \theta_i) \int \left\{ \frac{d}{dx} J[F(x)] \right\} dF(x)$$

and variance $\mathbf{b}^T \Upsilon \mathbf{b}$, where Υ is the $d \times d$ matrix

$$(\Upsilon)_{(ij)(km)} = \begin{cases} Q \lambda_i \lambda_j (\lambda_i + \lambda_j) & i = k, j = m, \\ Q \lambda_i \lambda_j \lambda_m & i = k, j \neq m, \\ Q \lambda_i \lambda_j \lambda_k & i \neq k, j = m, \\ Q \lambda_i \lambda_j \lambda_m & j = k \\ 0 & \text{otherwise,} \end{cases}$$

with $Q = \int_0^1 J^2(x) dx - \left[\int_0^1 J(x) dx \right]^2$.

Proceeding as in the proof of Corollary 3.1 we have the following.

Corollary 3.2 *Let F be a continuous distribution function with unique median η and strictly positive, continuous density $f(x)$ in some neighborhood of η . Further, assume the functions $J_{n_i+n_j}$ and F satisfy conditions B1 - B6. Then the asymptotic relative efficiency of the test V relative to the test W is given by*

$$e_{V,W}(F) = \frac{4f^2(\eta) \left\{ \int_0^1 J^2(x) dx - \left[\int_0^1 J(x) dx \right]^2 \right\}}{\left(\int_{-\infty}^{+\infty} \left\{ \frac{d}{dx} J[F(x)] \right\} dF(x) \right)^2}.$$

The efficiencies $e_{V,T}$ and $e_{V,W}$ are the same as found by Chernoff and Savage (1958) for the corresponding procedures in the two-sample problem, and shown

by Puri (1964) to be valid also for the multi-sample problem with unrestricted alternatives.

3.3 Equivalent statistics

Each test based on the normal theory statistic $T_N = \sum_{i < j} c_{ij} n_i n_j (\bar{X}_j - \bar{X}_i)$ can be written in the form $T_N = \sum_{i=1}^{c-1} c'_i n_i n_{i+1} (\bar{X}_{i+1} - \bar{X}_i)$, where the c'_i are functions of the c_{ij} and n_i . A natural question is whether a similar reduction holds for V_N . The purpose of this section is to develop the framework under which such an equivalence can be established. Having done so, it will be possible to derive tests that are “most efficient” against a given sequence of alternatives.

Let U_N and U_N^* be two statistics for testing H_0 versus H_1 . Tryon and Hettmansperger (1973) define U_N and U_N^* to be *equivalent* provided

1. Under H_0 , $U_N^* - U_N \xrightarrow{p} 0$, and
2. The asymptotic relative efficiency, $e_{U, U^*}(F)$, is unity for any choice of F in Ω .

We shall demonstrate that for each d -vector of coefficients (weights) \mathbf{b} , there exists a corresponding d -vector \mathbf{b}^* such that:

- $V_N(\mathbf{b}) = N^{-3/2} \mathbf{b}^T \mathbf{V}_N$ and $V_N(\mathbf{b}^*) = N^{-3/2} \mathbf{b}^{*T} \mathbf{V}_N$ are equivalent, and
- The elements of \mathbf{b}^* assign non-zero weight only to comparisons between adjacent groups, i.e.,

$$b_{ij}^* = 0, \quad j \neq i + 1, \quad i = 1, \dots, c - 1. \quad (3.11)$$

Another benefit of such a result is in reducing the number of pairwise tests to be

computed from $c(c-1)/2$ to $c-1$, a considerable improvement for even moderately large c .

Lemma 3.2 *Let A be a $d \times c$ matrix with components*

$$(A)_{(ij)k} = a_{(ij)k} = \begin{cases} \lambda_j & i = k, \\ -\lambda_i & j = k, \\ 0 & \text{otherwise.} \end{cases}$$

Let $\mathbf{b} \neq \mathbf{0}$ be a fixed d -vector. Then there exists a d -vector \mathbf{b}^ such that (3.11) is satisfied and $[\mathbf{b}^T - \mathbf{b}^{*T}]A = \mathbf{0}$. The elements of \mathbf{b}^* are given by*

$$b_{k(k+1)}^* = \frac{1}{\lambda_k \lambda_{k+1}} \sum_{m=1}^k \lambda_m \left[\sum_{j>m} b_{mj} \lambda_j - \sum_{i<m} b_{im} \lambda_i \right], \quad k = 1, \dots, c-1. \quad (3.12)$$

(The structure of the matrix A is displayed in Figure 1.1, page 12.)

Proof: To solve the relation $[\mathbf{b}^T - \mathbf{b}^{*T}]A = \mathbf{0}$ requires the following identity to hold for each $k : 1 \leq k \leq c$:

$$\sum_{i=1}^{c-1} \sum_{j=i+1}^c [b_{ij} - b_{ij}^*] a_{(ij)k} = 0.$$

For fixed $k : 1 \leq k \leq c$ this is equivalent to

$$b_{k(k+1)}^* \lambda_{k+1} - b_{(k-1)k}^* \lambda_{k-1} = \sum_{j>k} b_{kj} \lambda_j - \sum_{i<k} b_{ik} \lambda_i. \quad (3.13)$$

Substitute (3.12) into the left-hand side of (3.13):

$$b_{k(k+1)}^* \lambda_{k+1} - b_{(k-1)k}^* \lambda_{k-1} = \lambda_{k+1} \left\{ \frac{1}{\lambda_k \lambda_{k+1}} \sum_{m=1}^k \lambda_m \left[\sum_{j>m} b_{mj} \lambda_j - \sum_{i<m} b_{im} \lambda_i \right] \right\}$$

$$-\lambda_{k-1} \left\{ \frac{1}{\lambda_{k-1}\lambda_k} \sum_{m=1}^{k-1} \lambda_m \left[\sum_{j>m} b_{mj}\lambda_j - \sum_{i<m} b_{im}\lambda_i \right] \right\}.$$

The two summations over m differ only by the index $m = k$ in the first of them; the expression reduces to verify that (3.13) holds. ■

We now exhibit a test based only on comparisons of adjacent treatment groups that is equivalent to the test $N^{-3/2}\mathbf{b}^T\mathbf{V}_N$.

Theorem 3.3 *For each non-zero d -vector \mathbf{b} and corresponding test in \mathcal{L} , there exists a vector \mathbf{b}^* , and corresponding test in \mathcal{L}^* , such that (3.11) holds and $N^{-3/2}\mathbf{b}^T\mathbf{V}_N$ and $N^{-3/2}\mathbf{b}^{*T}\mathbf{V}_N$ are equivalent.*

Proof: Assume that H_0 is true. Take A as in (1.15). Apply Lemma 3.2 to construct \mathbf{b}^* such that $[\mathbf{b}^T - \mathbf{b}^{*T}]A = 0$. Using the representation of (1.13) we have

$$V_N(\mathbf{b}) - V_N(\mathbf{b}^*) = [\mathbf{b}^T - \mathbf{b}^{*T}]N^{-3/2}\mathbf{V}_N =$$

$$[\mathbf{b}^T - \mathbf{b}^{*T}]A B_N \mathbf{S}_N + [\mathbf{b}^T - \mathbf{b}^{*T}](A_N - A) B_N \mathbf{S}_N + [\mathbf{b}^T - \mathbf{b}^{*T}] \mathbf{R}_N. \quad (3.14)$$

The terms of (3.14) are, respectively, identically 0, $o_p(1)$ and $o_p(1)$, therefore the expression converges in probability to 0. As a result, $V_N(\mathbf{b})$ and $V_N(\mathbf{b}^*)$ have identical limiting distributions under H_0 . (An alternate argument would demonstrate that $[\mathbf{b}^T - \mathbf{b}^{*T}]\Sigma[\mathbf{b} - \mathbf{b}^*] = 0$.)

It remains to be shown that the two statistics have an asymptotic relative efficiency of 1. This requires

$$\sum_{i=1}^{c-1} \sum_{j=i+1}^c b_{ij}\lambda_i\lambda_j(\theta_j - \theta_i) = \sum_{i=1}^{c-1} b_{i(i+1)}^*\lambda_i\lambda_{i+1}(\theta_{i+1} - \theta_i). \quad (3.15)$$

Fix $k : 1 \leq k \leq c$. The coefficient of θ_k on the right-hand side of (3.15) is given by

$$b_{(k-1)k}^* \lambda_{k-1} \lambda_k - b_{k(k+1)}^* \lambda_k \lambda_{k+1} = \lambda_k \left[b_{(k-1)k}^* \lambda_{k-1} - b_{k(k+1)}^* \lambda_{k+1} \right].$$

By (3.13)

$$\begin{aligned} \lambda_k \left[b_{(k-1)k}^* \lambda_{k-1} - b_{k(k+1)}^* \lambda_{k+1} \right] &= \lambda_k \left[\sum_{i < k} b_{ik} \lambda_i - \sum_{j > k} b_{kj} \lambda_j \right] \\ &= \sum_{i < k} b_{ik} \lambda_i \lambda_k - \sum_{j > k} b_{kj} \lambda_k \lambda_j. \end{aligned}$$

This is precisely the coefficient of θ_k in the left-hand side of (3.15). The statement is true for each k and any θ_i , hence the sums in (3.15) are identical. ■

3.4 Optimal weighting coefficients

In view of the previous section's developments, restrict attention to the class \mathcal{L}^* of statistics based on linear combinations of Mood statistics for adjacent samples. Let $v_{iN} = [(n_i + n_{i+1})M_{i(i+1)} - \mu_{iN}]$, where $\mu_{iN} = (n_i + n_{i+1})n_i F_i(\eta_{i(i+1)})$. Denote the $(c-1)$ -vector of v_{iN} by \mathbf{V}_N . \mathcal{L}^* consists of the statistics

$$V_N(\mathbf{b}) = N^{-3/2} \mathbf{b}^T \mathbf{V}_N,$$

where \mathbf{b} is now a $(c-1)$ -vector of coefficients.

Under H_0 the asymptotic distribution of $V_N(\mathbf{b})$ has been shown to be nor-

mal with mean 0 and variance $\mathbf{b}^T \Sigma \mathbf{b}$ where Σ is $(c-1) \times (c-1)$:

$$\sigma_{ij} = \begin{cases} \frac{\lambda_i \lambda_{i+1} (\lambda_i + \lambda_{i+1})}{4} & j = i, \\ -\frac{\lambda_{j-1} \lambda_j \lambda_{j+1}}{4} & j = i \pm 1, \\ 0 & \text{otherwise.} \end{cases}$$

Hence

$$\mathbf{b}^T \Sigma \mathbf{b} = \frac{1}{4} \sum_{i=1}^{c-1} b_i^2 \lambda_i \lambda_{i+1} (\lambda_i + \lambda_{i+1}) - \frac{1}{2} \sum_{i=1}^{c-2} b_i b_{i+1} \lambda_i \lambda_{i+1} \lambda_{i+2}.$$

Under the sequence of translation alternatives H_1^M the distribution of $V_N(\mathbf{b})$ is limiting normal with mean 0 and variance $\mathbf{b}^T \Sigma \mathbf{b}$. The efficacy is given by

$$e_V = \frac{\left[f(\eta) \sum_{i=1}^{c-1} \lambda_i \lambda_{i+1} (\theta_{i+1} - \theta_i) \right]^2}{\mathbf{b}^T \Sigma \mathbf{b}}.$$

(The efficiency of one statistic with respect to another is the ratio of the two efficacies.)

For $i = 1, \dots, c-1$, let $\delta_i = (\theta_{i+1} - \theta_i) / (\theta_c - \theta_1)$, with $\boldsymbol{\delta}$ the $(c-1)$ -vector of δ_i . We shall refer to the δ_i as the *relative spacings*. For the sequence of hypotheses H_1^M which specify $\theta_{iN} = \theta_i / \sqrt{M}$ the δ_i remain constant.

Consider testing a restricted alternate hypothesis, one in which the relative spacings δ_i are assumed known, i.e.,

$$H_0 : \boldsymbol{\delta} = \mathbf{0} \text{ vs. } H_2 : \boldsymbol{\delta} > \mathbf{0},$$

where, under H_2 , $\boldsymbol{\delta} > \mathbf{0}$ is fixed. It is now possible to select a weighting vector $\hat{\mathbf{b}}$ for

which the efficacy of the test is maximized. Reparameterize in terms of the δ_i :

$$e_V = \frac{f^2(\eta) (\theta_c - \theta_1)^2 \left(\sum_{i=1}^{c-1} b_i \lambda_i \lambda_{i+1} \delta_i \right)^2}{\mathbf{b}^T \Sigma \mathbf{b}}.$$

It is convenient to make another change in notation. Let Λ be the $(c-1) \times (c-1)$ diagonal matrix, $(\Lambda)_{ii} = \lambda_i \lambda_{i+1}$, and let Γ be the $(c-1) \times (c-1)$ matrix

$$\gamma_{ij} = \begin{cases} \frac{\lambda_i + \lambda_{i+1}}{4\lambda_i \lambda_{i+1}} & i = j, \\ -\frac{1}{4\lambda_j} & j = i \pm 1, \\ 0 & \text{otherwise.} \end{cases}$$

Then $\mathbf{b}^T \Sigma \mathbf{b} = \mathbf{b}^T \Lambda \Gamma \Lambda \mathbf{b}$, and

$$\left(\sum_{i=1}^{c-1} b_i \lambda_i \lambda_{i+1} \delta_i \right)^2 = \mathbf{b}^T \Lambda \boldsymbol{\delta} \boldsymbol{\delta}^T \Lambda \mathbf{b}.$$

Write e_V as follows:

$$e_V = f^2(\eta) (\theta_c - \theta_1)^2 \frac{\mathbf{b}^T (\Lambda \boldsymbol{\delta}) (\Lambda \boldsymbol{\delta})^T \mathbf{b}}{\mathbf{b}^T \Lambda \Gamma \Lambda \mathbf{b}} = f^2(\eta) (\theta_c - \theta_1)^2 \frac{(\mathbf{b}^T \Gamma \boldsymbol{\delta})^2}{\mathbf{b}^T (\Lambda \Gamma \Lambda) \mathbf{b}}.$$

Theorem 3.4 *A vector $\hat{\mathbf{b}}$ which maximizes e_V for a given vector $\Lambda \boldsymbol{\delta}$ is*

$$\hat{\mathbf{b}} = (\Lambda \Gamma \Lambda)^{-1} \Lambda \boldsymbol{\delta} = \Lambda^{-1} \Gamma^{-1} \boldsymbol{\delta}.$$

Proof: The result is a standard result in matrix theory (see Arnold (1981), page 339, Lemma 18.14). It is contingent on the non-singularity of the matrix $\Lambda \Gamma \Lambda$. That Λ is non-singular is clear from its diagonal structure. Let Γ^* be the $(c-1) \times (c-1)$

symmetric matrix given by

$$(\Gamma^*)_{(ij)} = \begin{cases} 4 \left(\sum_{k \leq i} \lambda_k \right) \left(\sum_{k > j} \lambda_k \right) & i \leq j, \\ 4 \left(\sum_{k > i} \lambda_k \right) \left(\sum_{k \leq j} \lambda_k \right) & i > j. \end{cases}$$

Direct multiplication shows that $\Gamma^* \Gamma = \Gamma \Gamma^* = \mathbf{I}$, hence Γ is invertible ($\Gamma^{-1} = \Gamma^*$); the condition is met. ■

Note that for any given vector of weights \mathbf{b} (including $\hat{\mathbf{b}}$) the efficacy is invariant under multiplication of \mathbf{b} by a constant. The efficacy of the optimal test is given by

$$\begin{aligned} e_V &= f^2(\eta) (\theta_c - \theta_1)^2 \frac{(\Lambda^{-1} \Gamma^{-1} \boldsymbol{\delta})^T (\Lambda \boldsymbol{\delta}) (\Lambda \boldsymbol{\delta})^T (\Lambda^{-1} \Gamma^{-1} \boldsymbol{\delta})}{(\Lambda^{-1} \Gamma^{-1} \boldsymbol{\delta})^T (\Lambda \Gamma \Lambda) (\Lambda^{-1} \Gamma^{-1} \boldsymbol{\delta})} \\ &= f^2(\eta) (\theta_c - \theta_1)^2 \boldsymbol{\delta}^T \Gamma^{-1} \boldsymbol{\delta}. \end{aligned}$$

The efficiency of the optimally weighted test relative to a test with weights \mathbf{b} is then

$$e_{\hat{\mathbf{b}}, \mathbf{b}} = \frac{(\boldsymbol{\delta}^T \Gamma^{-1} \boldsymbol{\delta}) (\mathbf{b}^T \Sigma \mathbf{b})}{\left(\sum_{i=1}^{c-1} b_i \lambda_i \lambda_{i+1} \delta_i \right)^2} = \frac{(\boldsymbol{\delta}^T \Gamma^{-1} \boldsymbol{\delta}) (\mathbf{b}^T \Sigma \mathbf{b})}{\mathbf{b}^T \Lambda \boldsymbol{\delta} \boldsymbol{\delta}^T \Lambda \mathbf{b}}. \quad (3.16)$$

For the equal sample size, equal relative spacings case the optimal vector is given by $(\hat{\mathbf{b}})_i = i(c-i), i = 1, \dots, c-1$.

3.5 Examples

Take $c = 3$. We adopt the convention of setting $\delta_1 = 1$, and scale the solution so that $\hat{b}_1 = 1$. For such a case the optimal weighting vector is given by

$$\hat{b}_2 = \frac{\lambda_1 + (\lambda_1 + \lambda_2) \delta_2}{(\lambda_2 + \lambda_3) + \lambda_3 \delta_2}.$$

For the equal sample size design the optimal vector is $\hat{\mathbf{b}}^T = [1, (1 + 2\delta_2)/(2 + \delta_2)]$. In the case $\delta_2 = 1$ (i.e., the relative differences between the three adjacent population locations are equal) the optimal vector assigns equal weight to each of the adjacent pairwise comparisons. This test is equivalent to the test in \mathcal{L} that assigns equal weight to each of the pairwise comparisons. (For equal sample sizes within groups, the optimal test for equal relative spacings is equivalent to the all-pairs, equal weights test. This is not the case when the sample sizes are unequal.)

Continuing with the example, suppose that $\lambda_1 = 1 - 2\alpha$, $\lambda_2 = \lambda_3 = \alpha$, where $0 < \alpha < 1/2$. Such a design corresponds to a situation in which the treatments are ordered *a priori*, with the first population being an experimental control. Experimental units are then allocated to the second and third groups in equal proportions. Assuming equal relative spacings, the optimal vector is given by $\hat{\mathbf{b}}^T = [1, 2/(3\alpha) - 1]$. Using (3.16) it is possible to compute the efficiency of the optimally weighted test relative to that which assigns all $b_{ij} = 1$:

$$e_{\hat{\mathbf{b}}, \mathbf{b}}(\alpha) = \frac{(5 - 9\alpha)(1 - 5\alpha + 8\alpha^2)}{(1 - \alpha)^2}.$$

A plot of $e(\alpha)$ is shown in Figure 3.1. At $\alpha = 0$, $e(\alpha) = 5$, from that point $e(\alpha)$ decreases to a minimum of 1 at $\alpha = 1/3$ (the equal sample sizes case). Thereafter $e(\alpha)$ rises, to a local maximum of $e(.451416) = 1.16207$, after which the function again descends to a value of 1 at $\alpha = .5$. $e(1/6) = 1.96$; for “near” alternatives the optimally weighted test requires approximately one-half the sample size as does the all-pairs equally weighted test in order to achieve the same power.

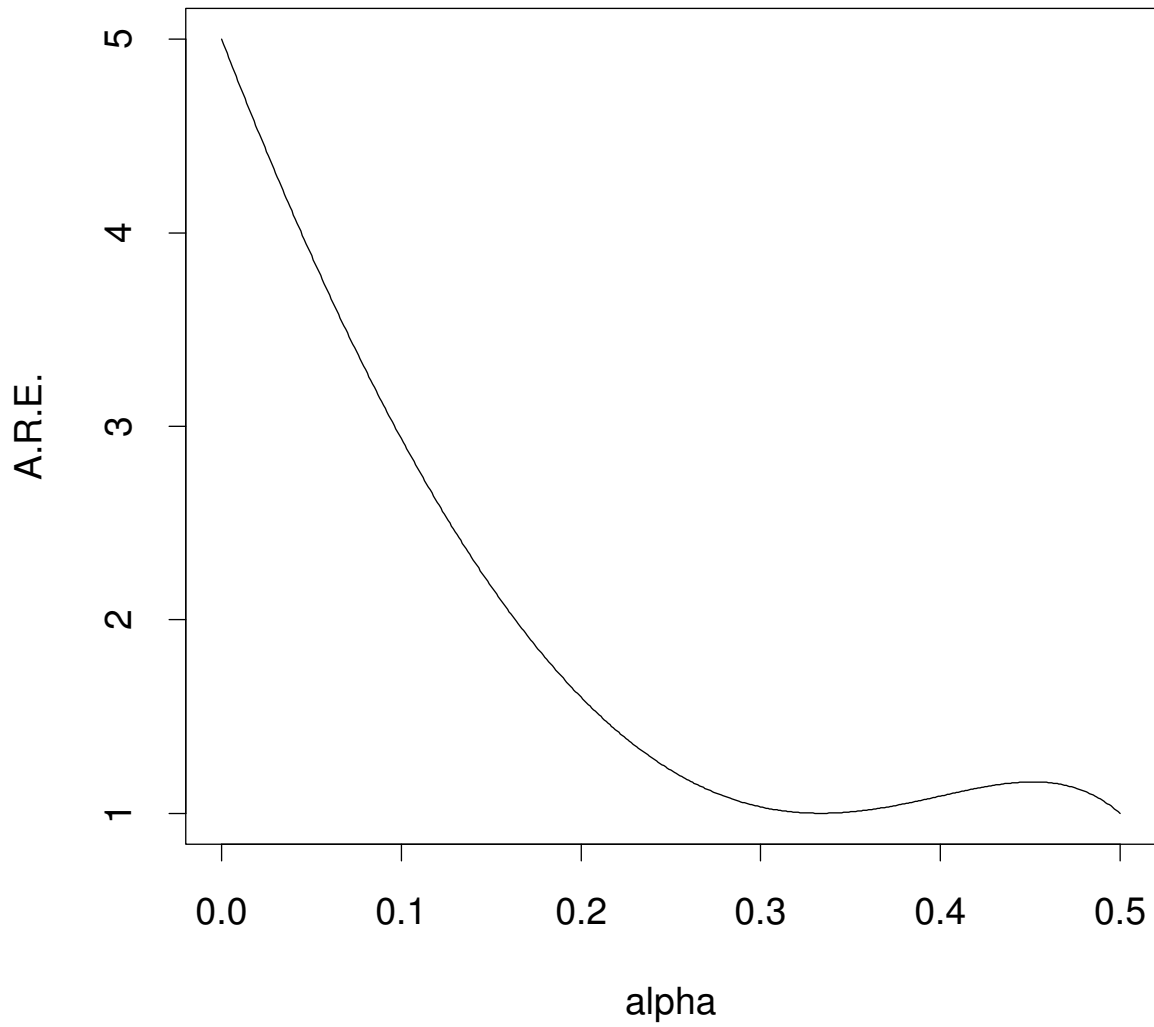


Figure 3.1: A.R.E. of Optimally Weighted Test Relative To All-Pairs, Equal Weights Test

3.6 Remarks

Tryon and Hettmansperger (1973) have proposed a weighting scheme for the pairwise adjacent Chernoff-Savage tests introduced early. The analysis is confined to equal sample sizes within groups. In such a case the weighting coefficients they derived are identical to those developed here.

Assuming a more general sampling scheme, it seems reasonable to expect that the results derived here apply to the test W_N based on the score function $J(u)$. As such, we state the following.

Corollary 3.3 *Assume that $J_{n_i+n_j}(u)$ satisfies conditions B1 - B6. Then*

1. *For each d -vector \mathbf{b} there exists a d -vector \mathbf{b}^* such that the statistics based on these two vectors are equivalent and \mathbf{b}_{ij}^* has non-zero components only for $j = i + 1$.*
2. *Restricted to the class of linear combinations of adjacent pairwise statistics and assuming hypothesized relative spacings, there exists an optimally weighted statistic. Further, the vector of optimal coefficients is computed in the same fashion as demonstrated in Theorem 3.4.*

The details of proof are omitted. The result applies largely because the form of the efficacy for the W test is similar to that of the V test. It has already been noted that $\Upsilon = 4Q\Sigma$.

Tryon and Hettmansperger conclude that the optimal test against equal relative spacings is fairly robust (in terms of lost efficiency) to various configurations of $\boldsymbol{\delta}$, the true vector of relative spacings. As the above example illustrates,

this robustness does not apply so generally to configurations of the sample proportions. This point has not been addressed in the literature. Robertson et al. (1988) comment:

If one believes the distributions are nearly “equally spaced,” then the Jonckheere-Terpstra test, or Puri’s modification of it using scores, should be used.

Outside of the equal sample sizes design this remark is not true. If it is desired to test, using these methods, with power concentrated at any vector of relative spacings, the methods demonstrated here will produce the proper coefficients.

The relative spacings against which a given test is optimal are given by $\boldsymbol{\delta} = \Gamma \Lambda \mathbf{b}$ (in the all-pairs case first compute the equivalent). Consider another version of the pairwise Mood test procedure,

$$V_N^* = N^{-1/2} \sum_{i=1}^{c-1} \sum_{j=i+1}^c (M_{ij} - n_i/2).$$

Taking $b_{ij} = (\lambda_i + \lambda_j)^{-1}$, the (“scored”) version studied in this chapter produces essentially the same test as V_N^* . The relative spacings against which this test is optimal are not equal to the relative spacings against which

$$V_N = N^{-1/2} \sum_{i=1}^{c-1} \sum_{j=i+1}^c (n_i + n_j) (M_{ij} - n_i/2).$$

is optimal. For $\boldsymbol{\lambda}^T = [(1 - 2\alpha), \alpha, \alpha]$, the test V_N is optimal against relative spacings $\boldsymbol{\delta}^T = [(1 - \alpha), 2\alpha]$, V_N^* is optimal against $\boldsymbol{\delta}^T = [(1 + \alpha), 2(1 - \alpha)]$.

As a final remark, a third sign-scored method for testing H_1 is based on the statistic

$$V_N^{**}(\mathbf{b}) = N^{-1/2} \sum_{i=1}^{c-1} \sum_{j=i+1}^c b_{ij} n_i n_j [\hat{F}_i(\hat{\eta}) - \hat{F}_j(\hat{\eta})],$$

where $\hat{\eta}$ is the aggregate sample median. V_N^{**} resembles the Mood statistic for the c -sample problem, unrestricted alternative. It can be verified that $V_N(\mathbf{b})$ and $V_N^{**}(\mathbf{b})$ are also equivalent. (The parametric test based on differences between medians is another equivalent.)

Chapter 4

Comparison of Procedures

This chapter begins with an asymptotic comparison of the statistics developed in Chapter 3. In testing an order restricted model it cannot be assumed that the relative spacings are known. What is of interest is the robustness (in terms of efficiency) of the relative spacings to mistaken assumptions on them. A variety of weighting schemes will be compared.

Following this we develop the asymptotic distribution of a competitor to the pairwise testing procedure, $\bar{\chi}_S^2$, the sign-scored analogue to the $\bar{\chi}^2$ test developed by Bartholomew (1959).

Lastly, the powers of various pairwise statistics and that of the $\bar{\chi}_S^2$ statistic will be simulated and compared for a collection of sample configurations and relative spacings.

4.1 Efficiency comparisons

Asymptotic relative efficiencies are presented for various weighting schemes. In each case a specified set of relative spacings has been selected. In practice, precise information on the relative spacings will not be available, although the experimenter may have some predisposition regarding the nature of the response. The tabled values show the effect that a mistaken assumption has on the efficiency of the test.

Four sample configurations have been selected: equal sample sizes, sample sizes descending (linearly) by treatment index, a “lightly” control dominated configuration (the control sample being twice the size of each of the others), and

a “heavily” control dominated configuration (the control sample being one-half the entire sample, with all other sample sizes equal).

The selected relative spacings are labeled A–H. Descriptions are given below. For the case $c = 10$ the response curves corresponding to these spacings have been plotted in Figures 4.1–4.8.

- A. Linear response: $\delta_i = 1, i = 1, \dots, c - 1$.
- B. Linear, with the exception that the spacing between the locations of the two “middle” treatment groups is double that of the other spacings: $\delta_{\lceil c/2 \rceil} = 2, \delta_i = 1, i \neq \lceil c/2 \rceil$.
- C. Linear, with the exception that the spacing between the locations of the first and second treatment groups is ten times that of the other spacings: $\delta_i = 1 + .1(i - 1), i = 1, \dots, c - 1$.
- D. Convex quadratic: $\delta_i = i, i = 1, \dots, c - 1$.
- E. Convex exponential: $\delta_i = 2^{i-1}, i = 1, \dots, c - 1$.
- F. Partial sums of the harmonic series: $\delta_i = 1/i, i = 1, \dots, c - 1$.
- G. Concave exponential: $\delta_i = 1/2^{i-1}, i = 1, \dots, c - 1$.
- H. Concave quadratic: $\delta_i = c - i + 1, i = 1, \dots, c - 1$.

In each case the following are computed:

e_1 : efficiency of the optimally weighted test relative to that which is optimal for the given sample configuration and relative spacings given by A,

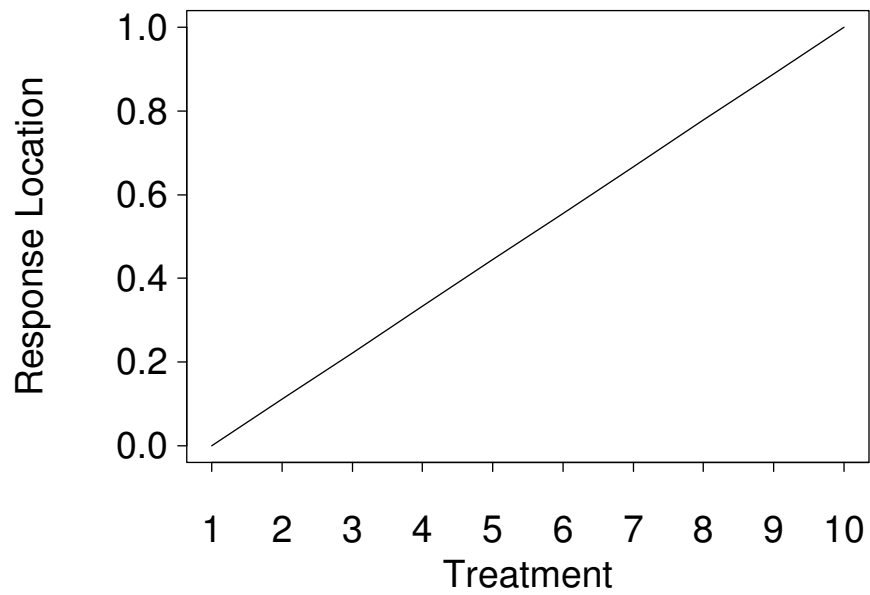


Figure 4.1: Response Curve A: Linear

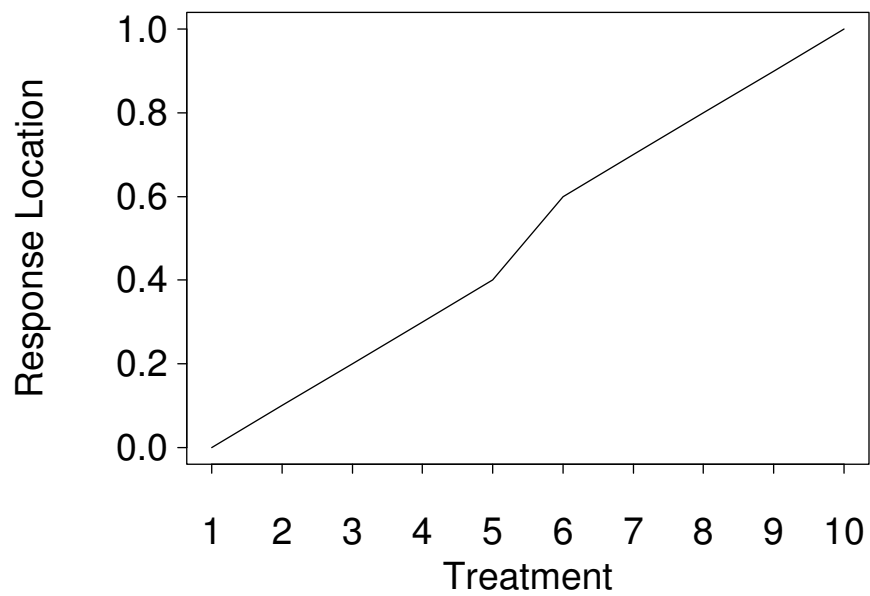


Figure 4.2: Response Curve B: Linear With Double Spacing Between the Middle Groups

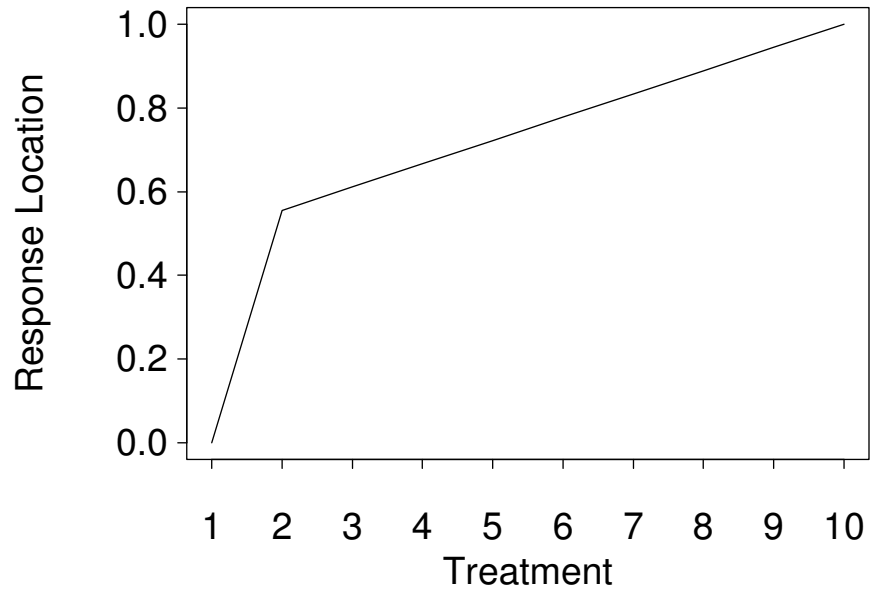


Figure 4.3: Response Curve C: Linear With Ten-fold Spacing Between the First Two Groups

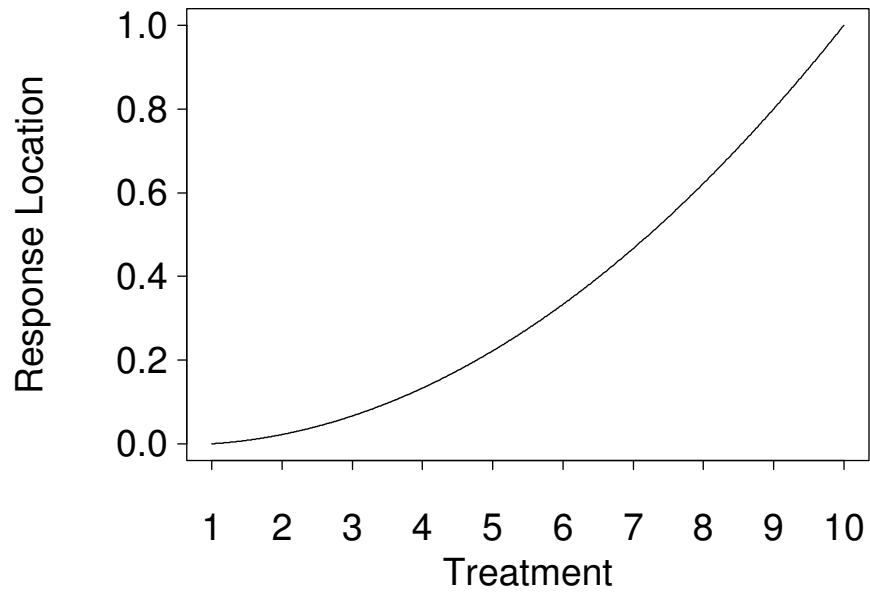


Figure 4.4: Response Curve D: Convex Quadratic

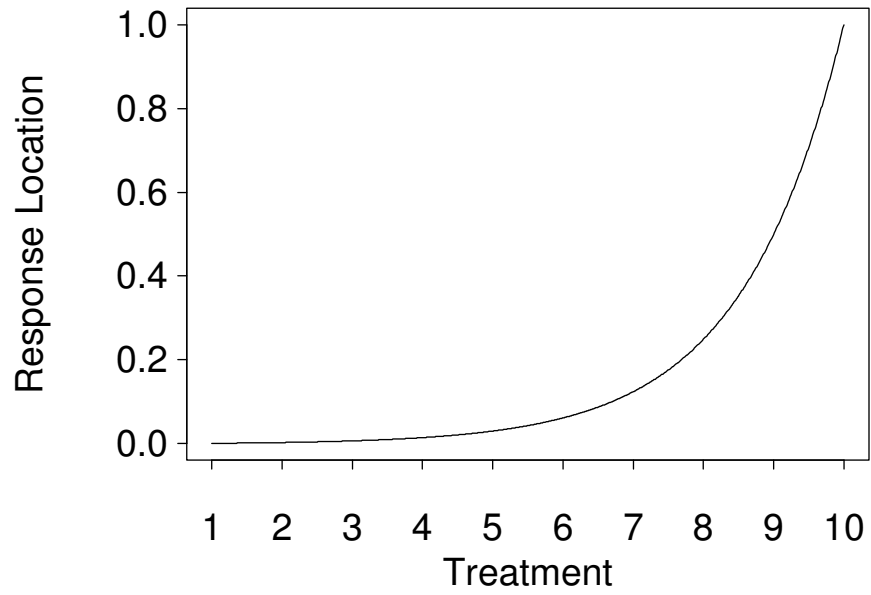


Figure 4.5: Response Curve E: Convex Exponential

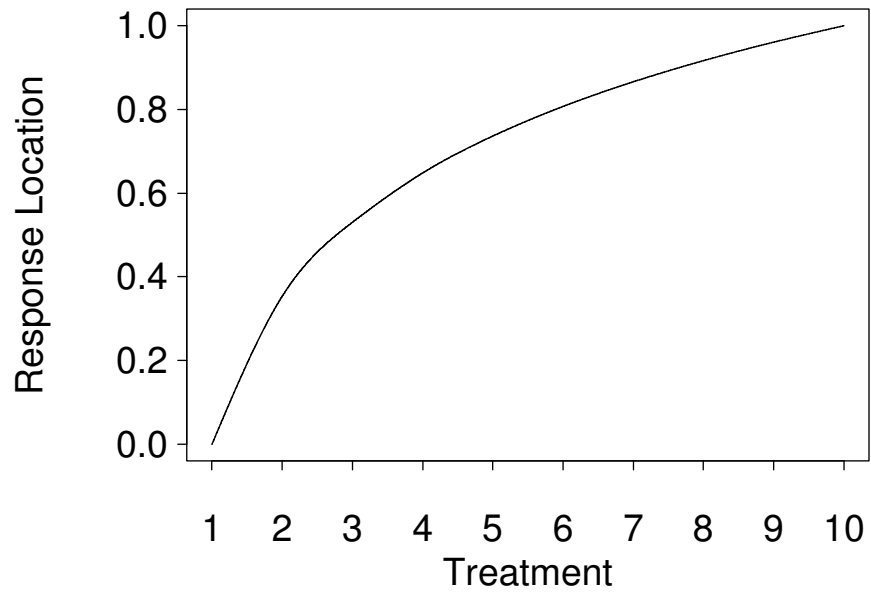


Figure 4.6: Response Curve F: Partial Sums of Harmonic Series

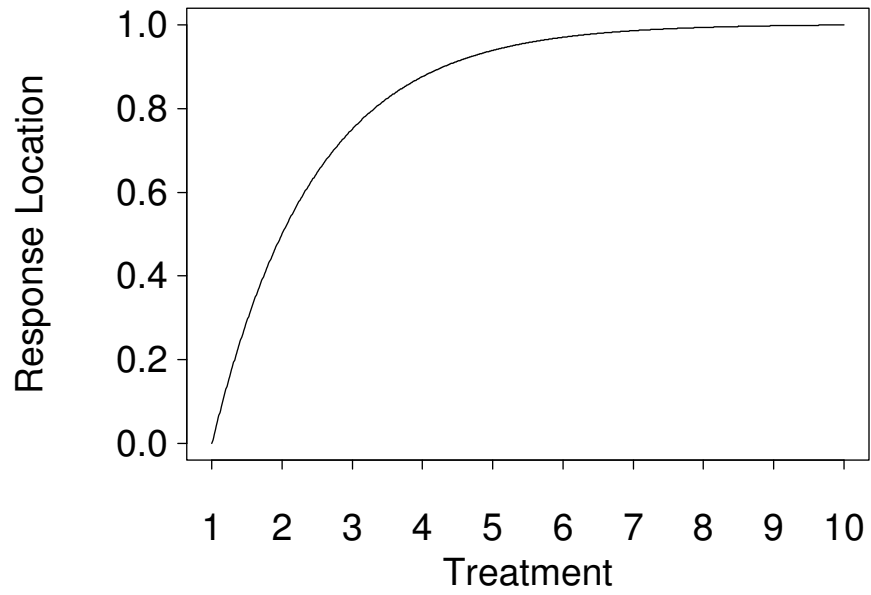


Figure 4.7: Response Curve G: Concave Exponential

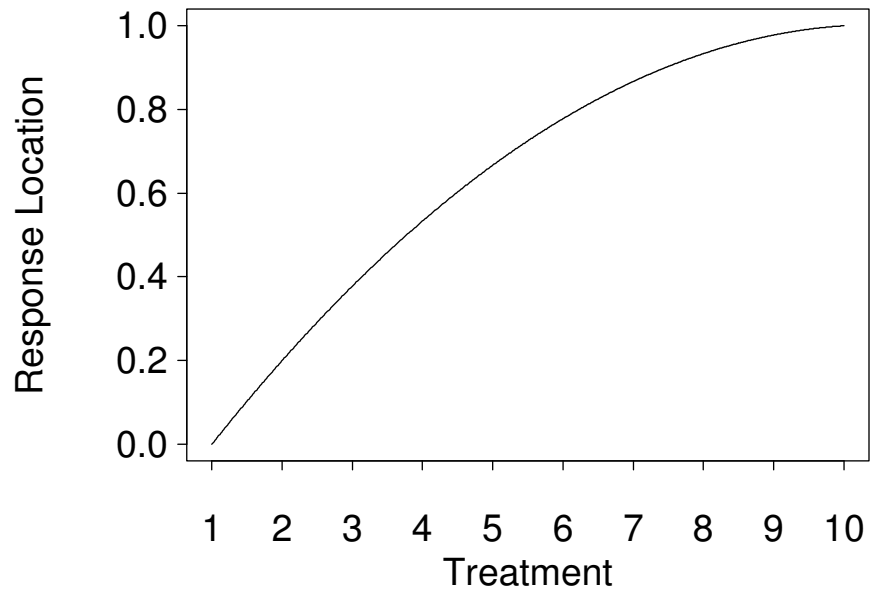


Figure 4.8: Response Curve H: Concave Quadratic

e_2 : efficiency of the optimally weighted test relative to that which is optimal for the equal sample size configuration (I) and given relative spacings, and

e_3 : efficiency of the optimally weighted test relative to the test given by all pairwise comparisons, equally weighted.

Results are tabulated for $c = 3, 4, 5, 7, 10$. The author has values for cases $c = 6, 8, 9$, as well as the optimal weighting coefficients for each combination of sample configuration and relative spacings.

The results in Table 4.1 agree with those for the equal sample-sizes pairwise rank tests given in Tryon and Hettmansperger (1973), who tabled values for the relative spacings given by A–E.

The efficiencies e_2 reinforce the remark made in Section 3.5. The optimal test for equal sample sizes and the given relative spacings (denoted \tilde{V}) is generally quite inefficient when sample sizes are unequal.

The test weighted for the given sample sizes and equal relative spacings (V^*) is quite robust against the violation of equal relative spacings. Only against the alternatives specifying an exponential response curve, where the spacings are doubling and halving, respectively, is the optimal statistic \hat{V} a significant improvement.

The all-pairs, equal weights test, V performs slightly better than does V^* for alternatives of the concave nature. Elsewhere, V^* is more efficient than is V . It appears that V is most powerful when the relative spacings and sample sizes increase (decrease) together.

These points are reinforced by the simulation presented in Section 4.3, where \hat{V} , V and V^* are compared with the sign-scored version of the $\bar{\chi}^2$ test.

Table 4.1: Asymptotic Relative Efficiencies, Equal Sample Sizes

c		Response Curve							
		A	B	C	D	E	F	G	H
3	$e_1 = e_3$	1.000	1.037	1.001	1.037	1.037	1.037	1.037	1.037
4	$e_1 = e_3$	1.000	1.020	1.002	1.050	1.087	1.065	1.087	1.050
5	$e_1 = e_3$	1.000	1.018	1.003	1.056	1.148	1.087	1.148	1.056
7	$e_1 = e_3$	1.000	1.010	1.005	1.061	1.298	1.119	1.298	1.061
10	$e_1 = e_3$	1.000	1.006	1.008	1.076	1.567	1.152	1.567	1.076

c	\longrightarrow	$[\lambda_1, \lambda_2, \dots, \lambda_c]$
3	\longrightarrow	all $\lambda_i = 1/3$
4	\longrightarrow	all $\lambda_i = 1/4$
5	\longrightarrow	all $\lambda_i = 1/5$
7	\longrightarrow	all $\lambda_i = 1/7$
10	\longrightarrow	all $\lambda_i = 1/10$

For equal sample sizes the efficiencies e_1 and e_3 are identical; the efficiency e_2 is identically 1 in all cases above.

Table 4.2: Asymptotic Relative Efficiencies, Descending Sample Sizes

		Response Curve							
c		A	B	C	D	E	F	G	H
3	e_1	1.000	1.046	1.001	1.046	1.046	1.035	1.035	1.035
	e_2	1.250	1.377	1.266	1.377	1.377	1.158	1.158	1.158
	e_3	1.020	1.136	1.030	1.136	1.136	1.002	1.002	1.002
4	e_1	1.000	1.026	1.002	1.065	1.110	1.063	1.078	1.043
	e_2	1.309	1.325	1.339	1.515	1.600	1.171	1.159	1.220
	e_3	1.029	1.072	1.046	1.197	1.280	1.008	1.012	1.001
5	e_1	1.000	1.023	1.003	1.074	1.193	1.086	1.128	1.046
	e_2	1.327	1.404	1.368	1.572	1.794	1.167	1.143	1.200
	e_3	1.033	1.098	1.057	1.229	1.448	1.015	1.030	1.001
7	e_1	1.000	1.014	1.005	1.083	1.416	1.122	1.244	1.047
	e_2	1.333	1.387	1.391	1.608	2.200	1.153	1.112	1.206
	e_3	1.037	1.087	1.072	1.260	1.878	1.031	1.086	1.001
10	e_1	1.000	1.008	1.009	1.088	1.892	1.160	1.444	1.096
	e_2	1.330	1.352	1.407	1.614	2.944	1.139	1.082	1.263
	e_3	1.039	1.060	1.089	1.277	2.800	1.050	1.198	1.048

c	\longrightarrow	$[\lambda_1, \lambda_2, \dots, \lambda_c]$
3	\longrightarrow	$[3, 2, 1] \div 6$
4	\longrightarrow	$[4, 3, 2, 1] \div 10$
5	\longrightarrow	$[5, 4, 3, 2, 1] \div 15$
7	\longrightarrow	$[7, 6, 5, 4, 3, 2, 1] \div 28$
10	\longrightarrow	$[10, 9, 8, 7, 6, 5, 4, 3, 2, 1] \div 55$

Table 4.3: Asymptotic Relative Efficiencies, Lightly Control Dominated

		Response Curve							
c		A	B	C	D	E	F	G	H
3	e_1	1.000	1.031	1.001	1.031	1.031	1.028	1.028	1.028
	e_2	1.222	1.222	1.224	1.222	1.222	1.195	1.195	1.327
	e_3	1.010	1.080	1.016	1.080	1.080	1.004	1.004	1.004
4	e_1	1.000	1.016	1.001	1.046	1.080	1.050	1.066	1.039
	e_2	1.288	1.283	1.280	1.224	1.201	1.296	1.295	1.304
	e_3	1.009	1.036	1.018	1.098	1.144	1.016	1.026	1.012
5	e_1	1.000	1.015	1.002	1.055	1.145	1.070	1.114	1.045
	e_2	1.296	1.260	1.278	1.198	1.150	1.350	1.355	1.351
	e_3	1.007	1.036	1.017	1.101	1.213	1.032	1.063	1.020
7	e_1	1.000	1.009	1.005	1.064	1.308	1.103	1.233	1.052
	e_2	1.269	1.239	1.237	1.153	1.081	1.398	1.423	1.362
	e_3	1.004	1.020	1.016	1.098	1.375	1.064	1.167	1.032
10	e_1	1.000	1.005	1.008	1.070	1.599	1.142	1.446	1.057
	e_2	1.220	1.204	1.180	1.113	1.038	1.410	1.477	1.322
	e_3	1.002	1.009	1.017	1.091	1.658	1.108	1.369	1.043

c	\longrightarrow	$[\lambda_1, \lambda_2, \dots, \lambda_c]$
3	\longrightarrow	$\lambda_1 = 1/2$, all other $\lambda_i = 1/4$
4	\longrightarrow	$\lambda_1 = 2/5$, all other $\lambda_i = 1/5$
5	\longrightarrow	$\lambda_1 = 1/3$, all other $\lambda_i = 1/6$
7	\longrightarrow	$\lambda_1 = 1/4$, all other $\lambda_i = 1/8$
10	\longrightarrow	$\lambda_1 = 2/11$, all other $\lambda_i = 1/11$

Table 4.4: Asymptotic Relative Efficiencies, Heavily Control Dominated

		Response Curve							
c		A	B	C	D	E	F	G	H
3	e_1	1.000	1.031	1.001	1.031	1.031	1.028	1.028	1.028
	e_2	1.222	1.222	1.224	1.222	1.222	1.195	1.195	1.327
	e_3	1.010	1.080	1.016	1.080	1.080	1.004	1.004	1.004
4	e_1	1.000	1.014	1.001	1.043	1.075	1.043	1.057	1.033
	e_2	1.938	2.034	1.942	1.888	1.817	1.795	1.783	1.852
	e_3	1.025	1.056	1.037	1.137	1.189	1.003	1.008	1.005
5	e_1	1.000	1.012	1.002	1.050	1.133	1.055	1.087	1.035
	e_2	3.145	3.112	3.143	2.935	2.526	2.732	2.666	2.986
	e_3	1.036	1.078	1.054	1.174	1.312	1.004	1.015	1.008
7	e_1	1.000	1.007	1.004	1.058	1.289	1.073	1.149	1.037
	e_2	6.983	6.825	6.835	5.817	3.515	5.483	5.052	6.891
	e_3	1.051	1.080	1.080	1.220	1.601	1.006	1.031	1.015
10	e_1	1.000	1.003	1.007	1.065	1.588	1.092	1.231	1.037
	e_2	15.828	15.625	14.647	11.136	3.952	11.483	9.549	16.700
	e_3	1.063	1.079	1.106	1.256	2.111	1.009	1.054	1.022

$c \longrightarrow [\lambda_1, \lambda_2, \dots, \lambda_c]$
 3 $\longrightarrow \lambda_1 = 1/2$, all other $\lambda_i = 1/4$
 4 $\longrightarrow \lambda_1 = 1/2$, all other $\lambda_i = 1/6$
 5 $\longrightarrow \lambda_1 = 1/2$, all other $\lambda_i = 1/8$
 7 $\longrightarrow \lambda_1 = 1/2$, all other $\lambda_i = 1/12$
 10 $\longrightarrow \lambda_1 = 1/2$, all other $\lambda_i = 1/18$

4.2 Sign-scored $\bar{\chi}^2$ test

Let $\hat{F}(\cdot)$ denote the aggregate sample distribution function and define $\hat{\eta}$ to be the median of the aggregate sample:

$$\hat{\eta} = \inf \{x : \hat{F}(x) \geq 1/2\},$$

Take $F_i^*(\hat{\eta})$ to be the antitonic regression of the $\hat{F}_i(\hat{\eta})$ with respect to the weight vector $\mathbf{w}_N^T = [n_1/N, \dots, n_c/N]$. The antitonic regression can be computed by the pool adjacent violators algorithm (PAVA), and is the projection of the (ordered) vector of $\hat{F}_i(\hat{\eta})$ onto the closed, convex cone of vectors antitonic with respect to the simple order and weights \mathbf{w}_N . The projection operator is continuous with respect to both its weights and its argument (Robertson et al. (1988), page 205). Then the sign-scored analogue to the $\bar{\chi}^2$ test for testing H_1 is given by

$$\bar{\chi}_S^2 = \sum_{i=1}^c 4n_i [F_i^*(\hat{\eta}) - 1/2]^2.$$

Let \mathbf{Y} be a vector of independent normal random variables, $Y_i \sim n(0, \lambda_i^{-1})$, $\bar{Y} = \sum_{i=1}^c \lambda_i Y_i$. Denote by $P_{\boldsymbol{\lambda}}(\mathbf{Y} - \bar{Y}\mathbf{1}|\mathcal{I})$ the projection of $[\mathbf{Y} - \bar{Y}\mathbf{1}]$ onto the set of c -vectors $\mathcal{I} = \{\mathbf{a} : a_1 \geq \dots \geq a_c\}$, with weights given by the c -vector of λ_i , denoted $\boldsymbol{\lambda}$. $P_{\boldsymbol{\lambda}}(\mathbf{Y} - \bar{Y}\mathbf{1}|\mathcal{I})_i$ is defined to be the i th component of the projection; a standard result is that $P_{\boldsymbol{\lambda}}(\mathbf{Y} - \bar{Y}\mathbf{1}|\mathcal{I}) = P_{\boldsymbol{\lambda}}(\mathbf{Y}|\mathcal{I}) - \bar{Y}\mathbf{1}$.

Theorem 4.1 *Assume $F(\cdot)$ is continuous at $\eta : F(\eta) = 1/2$.*

1. Under H_0

$$\bar{\chi}_S^2 \xrightarrow{\mathcal{L}} \|P_{\boldsymbol{\lambda}}(\mathbf{Y}|\mathcal{I}) - \bar{Y}\mathbf{1}\|_{\boldsymbol{\lambda}}^2 = \sum_{i=1}^c \lambda_i [P_{\boldsymbol{\lambda}}(\mathbf{Y} - \bar{Y}|\mathcal{I})_i]^2.$$

2. Let \bar{X}_i , $i = 1, \dots, c$ denote the subsample means and take \bar{X} to be the aggregate sample mean. Let X_i^* denote the isotonic regression of the \bar{X}_i . Define

$$\bar{\chi}_T^2 = \frac{1}{\sigma^2} \sum_{i=1}^c n_i (X_i^* - \bar{X})^2.$$

Then, under the sequence of alternatives H_1^M , further assuming $F(\cdot)$ to have a continuous density in some neighborhood of η and finite variance σ^2 , the asymptotic relative efficiency of the test $\bar{\chi}_S^2$ relative to $\bar{\chi}_T^2$ is given by

$$e_{\bar{\chi}_S^2, \bar{\chi}_T^2}(F) = 4f^2(\eta)\sigma^2.$$

Proof: Assume the existence of $\eta^\circ : \sum_{i=1}^c \lambda_i F_i(\eta^\circ) = 1/2$, where, for all i , $F_i(\cdot)$ has a continuous density $f_i(\cdot)$ in some neighborhood of η° . Define

$$\eta_N : \sum_{i=1}^c \frac{n_i}{N} F_i(\eta_N) = 1/2, \quad C = \sum_{i=1}^c \lambda_i f_i(\eta^\circ).$$

Using methods demonstrated in Chapter 2 write

$$\mathbf{U}_N = A_N \mathbf{S}_N + \mathbf{R}_N. \tag{4.1}$$

where

- \mathbf{U}_N is the c -vector of $2N^{1/2}[\hat{F}_i(\hat{\eta}) - F_i(\eta_N)]$,
- \mathbf{S}_N is the c -vector of $2n_i^{1/2}[\hat{F}_i(\eta_N) - F_i(\eta_N)]$,
- A_N is the $c \times c$ matrix

$$(A_N)_{ij} = \begin{cases} \left(1 - \frac{f_i(\eta^\circ)n_i}{CN}\right) \left(\frac{N}{n_i}\right)^{1/2} & i = j, \\ -\frac{f_i(\eta^\circ)n_j}{CN} \left(\frac{N}{n_j}\right)^{1/2} & \text{otherwise,} \end{cases}$$

- \mathbf{R}_N is a c -vector of terms $o_p(1)$.

Under H_0 , \mathbf{U}_N is limiting normal with mean vector $\mathbf{0}$ and covariance matrix Σ with elements $\sigma_{ii} = (1 - \lambda_i)/\lambda_i$, and $\sigma_{ij} = -1, i \neq j$. This is precisely the distribution of $[\mathbf{Y} - \bar{Y}\mathbf{1}]$.

Write

$$\bar{\chi}_S^2 = \sum_{i=1}^c \frac{n_i}{N} \left\{ P_{\mathbf{w}_N} \left[2N^{1/2} \left(\hat{F}_i(\hat{\eta}) - 1/2 \right) | \mathcal{I} \right]_i \right\}^2 = \sum_{i=1}^c \frac{n_i}{N} \left\{ P_{\mathbf{w}_N} [\mathbf{U} | \mathcal{I}]_i \right\}^2.$$

Appealing to the continuity of the projection,

$$\bar{\chi}_S^2 = \sum_{i=1}^c \frac{n_i}{N} \left\{ P_{\mathbf{w}_N} [\mathbf{U} | \mathcal{I}]_i \right\}^2 \xrightarrow{\mathcal{L}} \sum_{i=1}^c \lambda_i \left[P_{\boldsymbol{\lambda}} \left(\mathbf{Y} - \bar{Y}\mathbf{1} | \mathcal{I} \right)_i \right]^2 = \| P_{\boldsymbol{\lambda}} \left(\mathbf{Y} | \mathcal{I} \right) - \bar{Y}\mathbf{1} \|_{\boldsymbol{\lambda}}^2.$$

The random variable $\| P_{\boldsymbol{\lambda}} \left(\mathbf{Y} | \mathcal{I} \right) - \bar{Y}\mathbf{1} \|_{\boldsymbol{\lambda}}^2$ is distributed as the $\bar{\chi}^2$ random variable, therefore

$$\lim_{N \rightarrow \infty} P \left[\bar{\chi}_S^2 \geq a \right] = \sum_{i=1}^c P(i, k, \boldsymbol{\lambda}) P \left[\chi_{i-1}^2 \geq a \right],$$

where the $P(i, k, \boldsymbol{\lambda})$ are the level set probabilities (given weight vector $\boldsymbol{\lambda}$) and χ_ν^2 denotes a Chi-square random variable with ν degrees of freedom ($\chi_0^2 \equiv 0$).

The efficiency result is now verified. Define the c -vectors $\mathbf{U}_N(\mathbf{0})$, $\mathbf{U}_N(\boldsymbol{\theta})$ and $\Delta_N(\boldsymbol{\theta})$ as follows:

$$(\mathbf{U}_N(\mathbf{0}))_i = 2N^{-1/2} [\hat{F}_i(\hat{\eta}) - 1/2],$$

$$(\mathbf{U}_N(\boldsymbol{\theta}))_i = 2N^{-1/2} [\hat{F}_i(\hat{\eta}) - F_i(\eta_N)],$$

$$(\Delta_N(\boldsymbol{\theta}))_i = 2N^{-1/2} [F_i(\eta_N) - 1/2].$$

Then, using (4.1),

$$\mathbf{U}_N(\mathbf{0}) = \mathbf{U}_N(\boldsymbol{\theta}) + \Delta_N(\boldsymbol{\theta}) = A_N \mathbf{S}_N + \Delta_N(\boldsymbol{\theta}) + \mathbf{R}_N,$$

where \mathbf{R}_n is a c -vector of terms $o_p(1)$, and, for each $N(M)$, η_N and η° are computed under the assumption that H_1^M is the true hypothesis, $N = N(M)$ with $N/M \rightarrow \rho_S$.

As $M \rightarrow \infty$, \mathbf{S}_N converges in law to a c -variate normal random vector \mathbf{M} with $\mathbf{0}$ mean and identity covariance matrix. $A_N \rightarrow A$ where

$$(A)_{ij} = \begin{cases} (1 - \lambda_i) \lambda_i^{-1/2} & i = j, \\ \lambda_j \lambda_i^{-1/2} & i \neq j. \end{cases}$$

The vector $\Delta_N(\boldsymbol{\theta})$ converges to the vector Δ having i th component

$$(\Delta)_i = \lim_{M \rightarrow \infty} (\Delta_N(\boldsymbol{\theta}))_i = \lim_{M \rightarrow \infty} 2N^{1/2} [F_i(\eta_N) - 1/2] = (\rho_S)^{1/2} 2f(\eta) (\theta_i - \bar{\theta}),$$

where $\bar{\theta} = \sum_{i=1}^c \theta_i$; the details being essentially the same as those provided in Chapter 3. $\mathbf{R}_N \xrightarrow{p} 0$, hence the limiting distribution of $\mathbf{U}_N(\mathbf{0})$ is normal with mean vector Δ and covariance matrix Σ .

Let \mathbf{T}_N be the random vector having i th component $N^{1/2}(\bar{X}_i - \bar{X})/\sigma$, where \bar{X} is the aggregate sample mean. Under the sequence H_1^M , with $N'/M \rightarrow \rho_T$, it can be shown that the limiting distribution of the vector \mathbf{T}_N is normal with mean vector Δ' and covariance matrix Σ where

$$(\Delta')_i = \frac{(\rho_T)^{1/2} (\theta_i - \bar{\theta})}{\sigma}.$$

Again appealing to the continuity of the projection operator, as $M \rightarrow \infty$ the two tests will have equal power at the same sequence of alternatives when $\Delta = \Delta'$. The

result follows. ■

The asymptotic relative efficiency of $\bar{\chi}_S^2$ with respect to the Chernoff-Savage analogue to the $\bar{\chi}^2$ test can be established in a similar fashion (see Robertson et al. (1988), pages 207–208), and is identical to that given in Corollary 3.2.

4.3 Simulations

A simulation study was performed to compare the pairwise Mood tests to the $\bar{\chi}_S^2$ test. The selected sample configurations are identical to those chosen for the efficiency comparisons (see Figures 4.1–4.4). We assume that $n_i/N = \lambda_i$. The selected response curves follow. (The lettering scheme is described on page 55.)

- Linear (A).
- Convex quadratic (D).
- Concave quadratic (H).
- Convex exponential (E).
- Concave exponential (G).

The response curves corresponding to these relative spacings are displayed in Figure 4.9 for the case $c = 10$ (see also Figures 4.1, 4.4, 4.5, 4.7 and 4.8).

4.3.1 Design

The simulation was written in S-PLUS. Fortran code for weighted isotonic regression written by Ryan (1990) was translated into an S-PLUS function to compute an antitonic regression. For each combination of sample configuration

Response Curves: 10 Sample Design

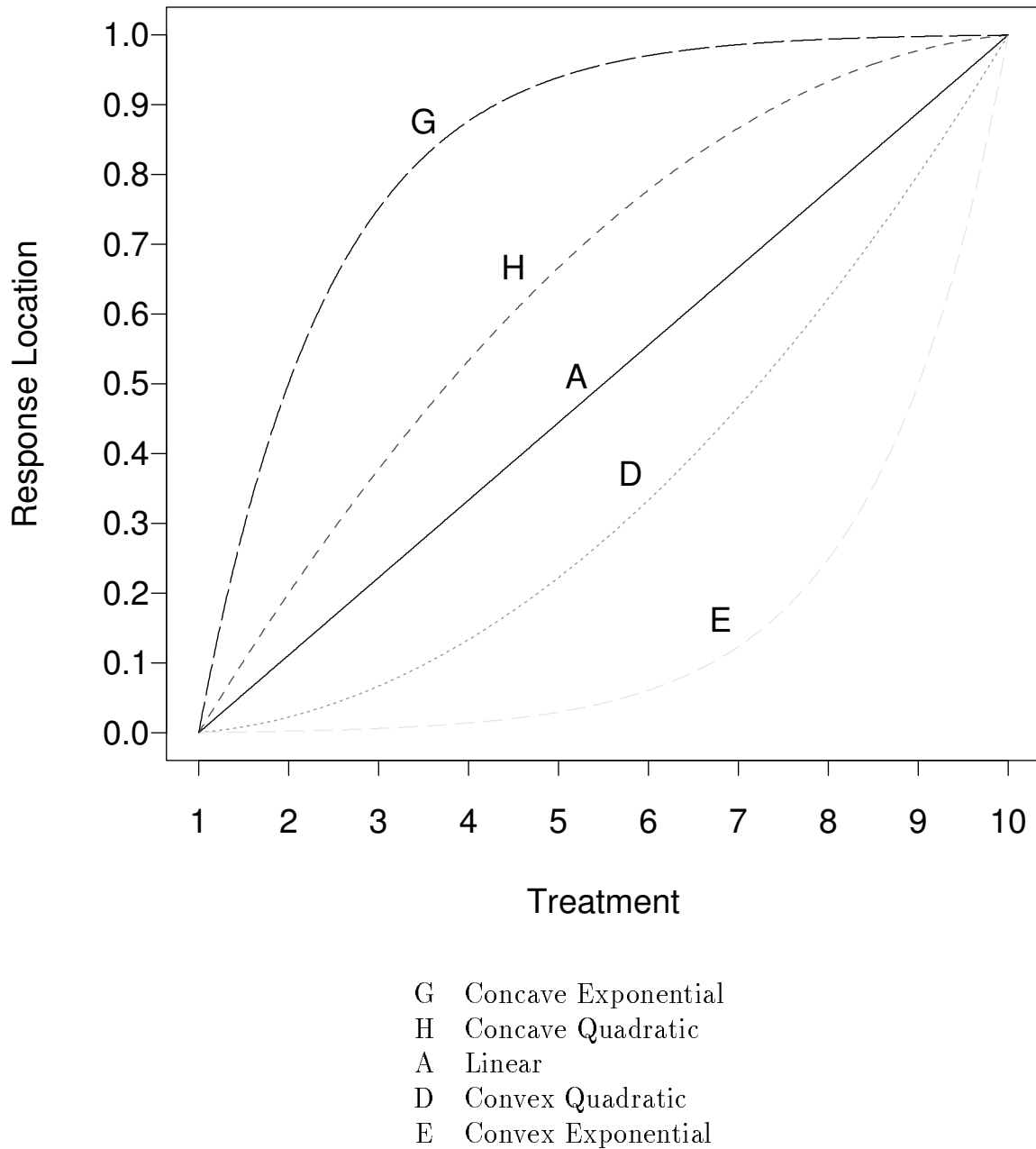


Figure 4.9: Response Curves Used in the Simulation

and relative spacings 10,000 samples were drawn. The significance level was set at $\alpha = .05$. Preliminary simulations were used to select location shifts achieving power centered about 0.5. The standard error of each value for simulated power is then no greater than 0.05.

The following statistics were considered:

\hat{V} : The adjacent pairs test, optimally weighted for the relative spacings as selected,

V : The equivalent to the all-pairs, equal weights test,

V^* : The adjacent pairs test, optimally weighted for equal relative spacings, and

$\bar{\chi}_S^2$: The sign-scored version of the $\bar{\chi}^2$ test.

The statistic optimally weighted against equal sample sizes and the given relative spacings (\tilde{V}) was not computed; efficiency results indicate that this test is generally inferior to \hat{V} , V , and V^* . The statistics \hat{V} , V , and V^* were implemented with a continuity correction. Calibrations under H_0 for a selected number of these tests proved quite accurate.

For the equal sample sizes case, critical values for the $\bar{\chi}^2$ statistic are tabled. For unequal sample sizes and $c \leq 4$, exact critical values were computed. Calibrations under H_0 were run to establish the integrity of these critical values. A slight conservatism was noted in some cases. In those cases the significance level for the tests \hat{V} , V , and V^* were set at the simulated levels (10,000 samples) of the $\bar{\chi}_S^2$ test, values ranging from 0.0434 to 0.0462.

For unequal samples sizes and $c \geq 5$ the situation is more difficult. The exact level probabilities are intractable. In the case of increased precision in a

control (configurations III and IV), the recommended approximation was developed by Chase (1974). The approach is to compute the $\bar{\chi}^2$ statistic twice, first using the level probabilities for the equal sample sizes case, then using limiting values for the level probabilities for the case in which the control sample mean is assumed fixed at the true mean. To compute a p -value Chase recommends interpolating on the two obtained p -values, with the interpolation on $w^{-1/2}$ where $w = (\lambda_1/\lambda_2)$. This approach was studied by Robertson and Wright (1982) and was found quite accurate. For this study, each approximation was calibrated with 10,000 samples under H_0 , the resulting simulated levels for these approximate tests ranged from 0.0474 to 0.0532. These simulated levels are within 1.5 standard deviations of the expected value of 0.05. As such, in this setting the significance levels for \hat{V} , V , and V^* were set at $\alpha = 0.05$.

For sample configuration IV an approximation based on the pattern of large and small weights, studied and recommended by Robertson and Wright (1983), was considered. The calibration procedure revealed a serious conservatism in results (simulated significance levels ranging from 0.01 to 0.03). The statistic was therefore simulated for 20,000 samples under H_0 , and the critical values were set at the 0.95 quantile of the simulated values. A comparison of the approximated and simulated critical values was not made; the method of Robertson and Wright relies on an approximation to the level probabilities that only implicitly defines the critical value.

4.3.2 Results

The adjacent pairs test, optimally weighted for (assumed) known relative spacings (\hat{V}) is superior to its competitors, particularly for larger c and against the more severe examples of increasing the relative spacings. If information on the

relative spacings is available, or can be hypothesized, \hat{V} is recommended.

The optimal test against equal relative spacings, V^* , is nearly as powerful as \hat{V} against the quadratic response curves. \hat{V} outperforms $\bar{\chi}_S^2$ against the linear and quadratic response curves. Against the convex exponential response curve and for $c \geq 5$, $\bar{\chi}_S^2$ is superior to V^* . This result is expected, as $\bar{\chi}_S^2$ is protecting against the entire response region. For control dominated samples (Tables 4.7 and 4.8), V^* compares favorably with $\bar{\chi}_S^2$ against the concave exponential response curve.

Efficiency results indicate that V , the equivalent to the all-pairs, equal weights test, is generally inferior to V^* except against alternatives in which the relative spacings and sample sizes decrease (increase) together. The simulated values reflect this (see Table 4.6). V was generally less powerful than its three competitors—faring better than V^* only against the convex exponential response curve.

In summary, the adjacent pairs, optimally weighted test, \hat{V} , and the adjacent pairs, optimally weighted for equal relative spacings, V^* , are recommended when it is reasonable to assume something about the nature of the response curve. V^* is nearly as powerful as is \hat{V} , provided the increases in relative spacings are not sharp. V^* is superior to $\bar{\chi}_S^2$ when the assumption of equal relative spacings is not seriously violated.

To protect well against *all* non-decreasing response curves, the $\bar{\chi}_S^2$ test is both appropriate and effective. The statistic V^* is far less powerful than $\bar{\chi}_S^2$ when the response curve has flat regions and/or sharp jumps.

It seems reasonable to expect similar results for tests based on Chernoff-Savage statistics.

Table 4.5: Simulated Power, Equal Sample Sizes

(c, N)		Response Curve		
		Linear	Quadratic Convex	Exponential Convex
(3,48)	$\pi(\hat{V})$	0.54	0.47	0.47
	$\pi(V^*)$	0.54	0.52	0.52
	$\pi(\bar{\chi}_S^2)$	0.43	0.42	0.42
(4,48)	$\pi(\hat{V})$	0.50	0.47	0.52
	$\pi(V^*)$	0.50	0.47	0.52
	$\pi(\bar{\chi}_S^2)$	0.45	0.42	0.48
(5,60)	$\pi(\hat{V})$	0.50	0.54	0.52
	$\pi(V^*)$	0.50	0.53	0.45
	$\pi(\bar{\chi}_S^2)$	0.48	0.47	0.46
(7,70)	$\pi(\hat{V})$	0.52	0.53	0.55
	$\pi(V^*)$	0.52	0.52	0.47
	$\pi(\bar{\chi}_S^2)$	0.46	0.47	0.44
(10,100)	$\pi(\hat{V})$	0.53	0.54	0.58
	$\pi(V^*)$	0.53	0.52	0.43
	$\pi(\bar{\chi}_S^2)$	0.47	0.47	0.47

(c, N)	\longrightarrow	$[n_1, n_2, \dots, n_c]$
(3, 48)	\longrightarrow	[16, 16, 16]
(4, 48)	\longrightarrow	[12, 12, 12, 12]
(5, 60)	\longrightarrow	[12, 12, 12, 12, 12]
(7, 70)	\longrightarrow	[10, 10, 10, 10, 10, 10, 10]
(10, 100)	\longrightarrow	[10, 10, 10, 10, 10, 10, 10, 10, 10, 10]

For equal sample sizes the all-pairs test (V) and the adjacent pairs test optimally weighted for equal relative spacings (V^*) are equivalent. Power against the convex response curves is equal to the power against the concave response curves.

Table 4.6: Simulated Power, Descending Sample Sizes

(c, N)		Response Curve				
		Linear	Quadratic		Exponential	
			Convex	Concave	Convex	Concave
(3,48)	$\pi(\hat{V})$	0.52	0.52	0.45	0.52	0.45
	$\pi(V)$	0.53	0.50	0.47	0.50	0.47
	$\pi(V^*)$	0.52	0.50	0.46	0.50	0.46
	$\pi(\bar{\chi}_S^2)$	0.47	0.47	0.41	0.47	0.41
(4,50)	$\pi(\hat{V})$	0.52	0.57	0.57	0.56	0.60
	$\pi(V)$	0.53	0.50	0.56	0.48	0.57
	$\pi(V^*)$	0.52	0.53	0.53	0.52	0.54
	$\pi(\bar{\chi}_S^2)$	0.47	0.50	0.49	0.51	0.50
(5,75)	$\pi(\hat{V})$	0.52	0.46	0.48	0.51	0.50
	$\pi(V)$	0.50	0.40	0.49	0.40	0.50
	$\pi(V^*)$	0.52	0.45	0.48	0.46	0.47
	$\pi(\bar{\chi}_S^2)$	0.47	0.42	0.43	0.48	0.43
(7,140)	$\pi(\hat{V})$	0.50	0.55	0.47	0.63	0.53
	$\pi(V)$	0.48	0.46	0.48	0.38	0.50
	$\pi(V^*)$	0.50	0.53	0.47	0.49	0.46
	$\pi(\bar{\chi}_S^2)$	0.44	0.49	0.41	0.51	0.45
(10,275)	$\pi(\hat{V})$	0.50	0.54	0.50	0.60	0.52
	$\pi(V)$	0.50	0.43	0.22	0.44	0.50
	$\pi(V^*)$	0.50	0.50	0.47	0.31	0.36
	$\pi(\bar{\chi}_S^2)$	0.44	0.45	0.41	0.42	0.38

(c, N)	\longrightarrow	$[n_1, n_2, \dots, n_c]$
(3, 48)	\longrightarrow	[24, 16, 8]
(4, 50)	\longrightarrow	[20, 15, 10, 5]
(5, 75)	\longrightarrow	[25, 20, 15, 10, 5]
(7, 140)	\longrightarrow	[35, 30, 25, 20, 15, 10, 5]
(10, 275)	\longrightarrow	[50, 45, 40, 35, 30, 25, 20, 15, 10, 5]

Table 4.7: Simulated Power, Lightly Control Dominated

(c, N)		Response Curve				
		Linear	Quadratic		Exponential	
			Convex	Concave	Convex	Concave
(3,48)	$\pi(\hat{V})$	0.53	0.54	0.53	0.54	0.53
	$\pi(V)$	0.53	0.51	0.56	0.53	0.56
	$\pi(V^*)$	0.53	0.51	0.56	0.51	0.56
	$\pi(\bar{\chi}_S^2)$	0.49	0.50	0.51	0.51	0.51
(4,50)	$\pi(\hat{V})$	0.50	0.55	0.57	0.56	0.58
	$\pi(V)$	0.54	0.55	0.58	0.54	0.57
	$\pi(V^*)$	0.50	0.53	0.54	0.52	0.52
	$\pi(\bar{\chi}_S^2)$	0.47	0.51	0.50	0.52	0.48
(5,60)	$\pi(\hat{V})$	0.53	0.54	0.60	0.60	0.56
	$\pi(V)$	0.53	0.51	0.58	0.53	0.54
	$\pi(V^*)$	0.53	0.53	0.58	0.55	0.52
	$\pi(\bar{\chi}_S^2)$	0.47	0.48	0.52	0.51	0.48
(7,80)	$\pi(\hat{V})$	0.56	0.54	0.52	0.51	0.60
	$\pi(V)$	0.55	0.50	0.51	0.40	0.54
	$\pi(V^*)$	0.56	0.52	0.51	0.42	0.52
	$\pi(\bar{\chi}_S^2)$	0.50	0.48	0.46	0.41	0.51
(10,110)	$\pi(\hat{V})$	0.51	0.56	0.51	0.60	0.56
	$\pi(V)$	0.51	0.52	0.49	0.42	0.44
	$\pi(V^*)$	0.51	0.53	0.49	0.44	0.43
	$\pi(\bar{\chi}_S^2)$	0.45	0.49	0.43	0.49	0.43

(c, N)	\longrightarrow	$[n_1, n_2, \dots, n_c]$
(3, 48)	\longrightarrow	[24, 12, 12]
(4, 50)	\longrightarrow	[20, 10, 10, 10]
(5, 60)	\longrightarrow	[20, 10, 10, 10, 10]
(7, 80)	\longrightarrow	[20, 10, 10, 10, 10, 10, 10]
(10, 110)	\longrightarrow	[20, 10, 10, 10, 10, 10, 10, 10, 10, 10]

Table 4.8: Simulated Power, Heavily Control Dominated

		Response Curve				
		Linear	<u>Quadratic</u>		<u>Exponential</u>	
(c, N)			Convex	Concave	Convex	Concave
(4,48)	$\pi(\hat{V})$	0.53	0.52	0.53	0.53	0.55
	$\pi(V)$	0.49	0.44	0.52	0.44	0.50
	$\pi(V^*)$	0.53	0.50	0.55	0.52	0.52
	$\pi(\bar{\chi}_S^2)$	0.46	0.45	0.48	0.47	0.46
(5,80)	$\pi(\hat{V})$	0.50	0.50	0.47	0.54	0.53
	$\pi(V)$	0.46	0.45	0.48	0.45	0.53
	$\pi(V^*)$	0.50	0.50	0.49	0.52	0.52
	$\pi(\bar{\chi}_S^2)$	0.44	0.46	0.45	0.48	0.49
(7,120)	$\pi(\hat{V})$	0.50	0.56	0.53	0.59	0.55
	$\pi(V)$	0.49	0.49	0.52	0.43	0.53
	$\pi(V^*)$	0.50	0.53	0.50	0.49	0.48
	$\pi(\bar{\chi}_S^2)$	0.46	0.51	0.47	0.51	0.45
(10,180)	$\pi(\hat{V})$	0.58	0.48	0.52	0.53	0.53
	$\pi(V)$	0.57	0.41	0.52	0.32	0.51
	$\pi(V^*)$	0.58	0.45	0.50	0.38	0.45
	$\pi(\bar{\chi}_S^2)$	0.53	0.44	0.45	0.44	0.43

(c, N)	\longrightarrow	$[n_1, n_2, \dots, n_c]$
(4, 48)	\longrightarrow	[24, 8, 8, 8]
(5, 80)	\longrightarrow	[40, 10, 10, 10, 10]
(7, 120)	\longrightarrow	[60, 10, 10, 10, 10, 10, 10]
(10, 180)	\longrightarrow	[90, 10, 10, 10, 10, 10, 10, 10, 10, 10]

Chapter 5

Resistance to Rejection of the One and Two-Sample Median Tests

The robustness of hypothesis tests has been studied by a number of authors. Ylvisaker (1977) proposed a definition of test resistance in the univariate setting using an additive contamination model. He, Simpson, and Portnoy (1990) considered the breakdown functions of a test statistic in functional form, which is an asymptotic version of Ylvisaker's resistance. Coakley and Hettmansperger (1992b) considered a variation of Ylvisaker's resistance based on a replacement contamination model.

5.1 Motivation

To motivate a definition of robustness of a test we think of a researcher collecting data, and a contaminator (of malicious nature), who alters it in such a way as to force a particular outcome on the experiment. A complete sample is obtained; some of the observations may be contaminated and the researcher does not know which are good and which are bad.

The previous work has viewed resistance to rejection of a test statistic when the data most favor the null hypothesis. For example, Coakley and Hettmansperger (1992b) defined the *maximum resistance to rejection* of a test as the smallest amount of (replacement) contamination necessary to force rejection no matter what the rest of the data are. Maximum resistance to rejection gives the resistance for the sample for which it is "hardest" to force rejection. This sample is, of course, quite unlikely; for most samples it will take far fewer contaminations than this number. When the

test statistic is slightly less than the critical value it may be sufficient to alter but a single sample value to force rejection. As a result, a measure of robustness which averages over possible samples is preferred.

5.1.1 Expected resistance to rejection

Coakley and Hettmansperger (1992b) have proposed the *expected resistance to rejection* (ERR) of a test as a measure of the average case robustness of a test.

Suppose that T_n is a real valued test statistic defined on the combined sample S , c_n is a critical value and a test procedure is given by

$$\text{reject } H_0 \text{ if } T_n(S) \geq c_n, \text{ otherwise fail to reject } H_0.$$

S represents the combined sample of size n . Let S_r^* be any sample obtained by replacing r of the points in S with arbitrary values. Define the *expected resistance to rejection* (ERR) of (T_n, c_n) by

$$\text{ERR} (T_n, c_n) = E_0 [R_n] / n,$$

where for each sample S ,

$$R_n = \min \left\{ r : \sup_{S_r^*} T_n \geq c_n \right\}.$$

R_n is the minimum number of contaminated observations need to force rejection. $\text{ERR}(T_n, c_n)$ is the expected value (taken under the null) of this minimal contamination. (Under H_0 the test will reject with no contamination with probability α .)

For fixed $\alpha \in (0, 1)$ let $\{(c_n, \alpha_n)\}$ be a sequence of critical values and

significance levels such that

$$c_n = \inf \{c : P [T_n \geq c] \leq \alpha\}, \quad \alpha_n = P [T_n \geq c_n].$$

Coakley and Hettmansperger (1992b) have shown that the one sided sign test has an expected resistance to rejection of

$$\text{ERR} (S_n^+, c_n) = \frac{c_n - n/2}{n},$$

where $S_n^+ = \sum_{i=1}^n I(X_i > \eta)$ is the test statistic.

As $n \rightarrow \infty$, $\alpha_n \rightarrow \alpha$, and $(n/4)^{-1/2}(c_n - n/2) \rightarrow z^\alpha$,

$$\text{ERR} (S_n^+, c_n) \approx n^{-1/2} z^\alpha / 2 + o(n^{-1/2}),$$

which converges to 0 at a rate of $n^{-1/2}$ as $n \rightarrow \infty$ with $0 < \alpha < 1$ fixed.

As $n \rightarrow \infty$ the expected resistance to rejection goes to 0. It is possible to norm ERR, and thus get a non-zero limiting measure of expected resistance to rejection. Define

$$\text{ERR}^* (T_n, c_n) = \lim_{n \rightarrow \infty} E_0 [n^{-1/2} R_n].$$

In the case of S_n^+

$$\begin{aligned} \text{ERR}^* (S_n^+, c_n) &= \lim_{n \rightarrow \infty} E \left\{ n^{-1/2} (c_n - S_n^+) I (S_n^+ < c_n) \right\} \\ &= \frac{1}{2} \left\{ z^\alpha (1 - \alpha) + \frac{\exp [-(z^\alpha)^2 / 2]}{(2\pi)^{1/2}} \right\}. \end{aligned}$$

A proof of this is routine, and is omitted.

Exact and approximate values are shown in Table 5.1. The exact signifi-

cance level was selected to be conservative with respect to a nominal $\alpha = .05$. The values suggest that a continuity correction of approximately +0.4 is appropriate for the approximate counts.

Table 5.1: Expected Resistance to Rejection of the One-Sided One-Sample Sign Test, Exact and Approximate Values

n	α	<u>Exact</u>		<u>Approximate</u>	
		Count	Proportion	Count	Proportion
30	.02139	5.0330	.1678	5.5753	.1858
60	.04623	7.0516	.1175	6.5904	.1098
120	.04120	10.0710	.0839	9.6050	.0800

For reasonably small sample sizes $n^{1/2}ERR^*$ provides fairly good approximations to the expected breakdown counts of the sign test. This will also shown to be the case for the Wilcoxon signed rank procedure. Further, ERR^* can be used as a comparison between procedures; the *relative robustness* of two competing procedures can be evaluated as

$$RR(T_{1n}, T_{2n}) = \lim_{n \rightarrow \infty} \frac{ERR(T_{1n}, c_{1n})}{ERR(T_{2n}, c_{2n})} = \frac{ERR^*(T_{1n}, c_{1n})}{ERR^*(T_{2n}, c_{2n})},$$

provided $ERR^*(T_{2n}) > 0$. (In the case of the t -test, at most one contamination always suffices: $ERR^*(t_n, t^\alpha) = 0$.)

5.1.2 Extensions

To conclude the chapter we derive ERR^* for the Mood median statistic. The development is more difficult than that for the 1-sample sign test, as each

permutation of the 2 samples must be considered separately. Uniform convergence results will allow us to approximate the optimal contamination by the scheme allowing only the alteration of values from the smaller sample, or either sample in case of equal sample sizes. This also is the case for the breakdown measures cited above; asymptotically, control of one-half the smaller sample forces rejection against the remainder of data most favorably situated for H_0 .

Associated with the influence curve for an estimate $T(F_n)$ is a measure of “local shift sensitivity,”

$$\lambda^* = \sup_{x \neq y} \left| \frac{IC(F; x) - IC(F; y)}{x - y} \right|,$$

measuring the local effects of rounding or grouping (contaminating) the observations. For the median functional $T(F) = F^{-1}(1/2)$, $\lambda^* = \infty$. To conclude the development of the expected resistance of the Mood test, the practical nature of this flaw in the testing situation will be examined. It will be shown that a contamination scheme exists for which the sum change in value of all contaminated data is, with probability one, $O(\log n)$ (and $O_p(1)$) as $n \rightarrow \infty$. Asymptotically, the entire analysis takes place in an $O(n^{-1/2} (\log n)^{1/2})$ neighborhood of the true median.

In Chapter 6 the resistance of the one-sided Wilcoxon test, W_n , is examined. Our results indicate that $RR(W_n, S_n^+) = 1/\sqrt{3}$.

5.2 Two-sample setting, single sample contamination schemes

Assume $X_i : i = 1, \dots, n_x$, $Y_j : j = 1, \dots, n_y$ are *i.i.d.* samples distributed according to some distribution function F . As Mood’s test is distribution free, and to increase clarity, without loss of generality take F to be the uniform distribution

on $(-1/2, 1/2)$. Denote by \hat{F} the empirical distribution function of the aggregate sample, and by \hat{F}_x and \hat{F}_y respectively the empirical distribution functions of the X_i and Y_j . Let $n = n_x + n_y$ with $n_x/n = \lambda + O(n^{-1})$, $\lambda \in (0, 1)$. We focus on the two-sample Mood median test M , one-sided alternative $H_1 : \eta_x > \eta_y$. Define

$$\hat{\eta} = \inf \left\{ t : \hat{F}(t) \geq 1/2 \right\}, \quad M = n_x \left[1 - \hat{F}_x(\hat{\eta}) \right].$$

The test rejects when the number, M , of X_i greater than the combined sample median is sufficiently large. We begin by examining a contamination scheme that alters only values of the X_i .

One indication that the solution to this restricted problem might be near optimal comes from the representation of the median statistic given in Theorem 1.1:

$$\left(M - \frac{n_x}{2} \right) = \frac{n_y}{n} S_x(\eta) - \frac{n_x}{n} S_y(\eta) + o_p(n^{1/2}), \quad (5.1)$$

where $S_x(\eta) = \sum_{i=1}^{n_x} I(X_i > \eta)$ and $S_y(\eta) = \sum_{j=1}^{n_y} I(Y_j > \eta)$. Ignoring the remainder term, the statistic on the right hand side of (5.1) could be used to test

$$H_0 : \eta_x = \eta = \eta_y, \quad H_1 : \eta_x \leq \eta \leq \eta_y,$$

with η assumed known. Suppose $n_y > n_x$; then contaminating X_i 's (by taking data values below η and assigning them values greater than η) will be more expeditious in forcing rejection than contaminating Y_j 's, as the increase in the test statistic after a single contamination of an X_i is greater than the corresponding increase resulting from the contamination of a single Y_j .

The median test rejects when $M \geq c_n$, where c_n is the (integer-valued) α_n level critical value determined by the null distribution (hypergeometric) of M . As

$n \rightarrow \infty$

$$(\text{Var}[M])^{-1/2} \left(c_n - \frac{n_x}{2} \right) \rightarrow z^\alpha.$$

We assume that rejection is *attainable*, that is, if $M = \min\{n_x, n/2\}$ then $M \geq c_n$. This assures us that the test can be forced to reject. For any fixed α , as n increases ($n_x/n \rightarrow \lambda$) rejection will eventually be attainable. Denote by ERR_x^* the analogue to ERR^* with the contamination scheme is restricted to altering the X_i .

Let $D = c_n - M$, the difference between the test statistic and the critical value. If $D \leq 0$ no contamination is necessary to force rejection. For $D = d > 0$, since altering only X_i 's permitted, the scheme will take observations $X_i < \hat{\eta}$ and successively reassign them values greater than $\hat{\eta}$ —for now assume the replacement value puts a contaminated item as the largest value in the combined sample. Each time an X_i is contaminated as such, the overall sample median is “pushed” up and becomes the value adjacent, and above, the current sample median. In contaminating the first X_i , the value of the test statistic will be increased by 1 if the data item immediately adjacent, and above, $\hat{\eta}$ is a Y_j , otherwise the test statistic remains unaltered. The procedure iterates until the location of the sample median has been pushed through d Y_j 's, at which point rejection has been forced.

(It is possible to create samples for which the strategy does not appear to force rejection: consider $c_n = 2$ and ordered data $\{Y \ X \ Y \mid X \ Y \ Y\}$, where \mid represents the location of $\hat{\eta}$. After contaminating all (1) of the X_i below $\hat{\eta}$ we have $\{Y \ Y \ X \mid Y \ Y \ X\}$ and rejection has not been forced. But, as we have assumed that c_n is attainable, we can apply the scheme again—in general contaminating X_i that have been moved below the position of the up-dated sample median in the initial contamination—resulting in $\{Y \ Y \ Y \mid Y \ X \ X\}$, and forcing rejection in 2

contaminations. For large n this situation will occur rarely, but the contamination mechanism is thus extended to cover this possibility.)

Contaminating the smaller sample is preferable to contaminating the larger sample: the crucial point is not the nature (whether X_i or Y_j) of the altered data, but instead the nature of the match between the contaminated values and the values adjacent $\hat{\eta}$. A scheme works best when contaminated values tend to be from the smaller sample—there will be fewer matches and, hence, more progress toward rejection.

For fixed $D = d > 0$ (and hence $M = m < c_n$), the number of contaminations necessary to force rejection, R_n^x , is a negative hypergeometric random variable

$$f_{R_n^x|m}(r) = P[R_n^x = r | M = m] = \frac{\binom{r-1}{d-1} \binom{n/2-r}{n/2-c_n}}{\binom{n/2}{m}}, \quad r = d, \dots, c_n. \quad (5.2)$$

To derive this note that there are exactly m X_i and $(n/2 - m)$ Y_j above $\hat{\eta}$. Any permutation of these is equally likely. Then the probability that $[R_n^x = r]$ is the probability of exactly $(d-1)$ Y_j in the first $(r-1)$ observations above, and adjacent, $\hat{\eta}$, and a Y_j as the r th value above $\hat{\eta}$.

$$P[R_n^x = r | M = m] = \frac{\binom{n/2-m}{d-1} \binom{m}{r-d}}{\binom{n/2}{r-1}} \times \frac{n/2 - (d-1)}{n/2 - (r-1)},$$

which reduces to (5.2). A standard result on the negative hypergeometric gives

$$E [R_n^x | M = m] = \begin{cases} \frac{(n/2 + 1)(c_n - m)}{n/2 + 1 - m} & m < c_n, \\ 0 & m \geq c_n, \end{cases}$$

and ERR_x^* can be computed by taking an expectation with respect to D (i.e. M).

$$\begin{aligned} ERR_x^* &= \lim_{n \rightarrow \infty} E [n^{-1/2} R_n^x] = \lim_{n \rightarrow \infty} E_M \{ E [(n^{-1/2} R_n^x) | M] \} \\ &= \lim_{n \rightarrow \infty} E \left[\frac{(n/2 + 1)(c_n - M)}{n^{1/2}(n/2 + 1 - M)} I(M < c_n) \right]. \end{aligned}$$

Lemma 5.1 *Define*

$$U_n = \frac{(n/2 + 1)(c_n - M)}{n^{1/2}(n/2 + 1 - M)} I(M < c_n),$$

and denote by Z the standard normal random variable. Then

1. $U_n \xrightarrow{\mathcal{L}} U$ where U is a random variable having the distribution of

$$\frac{1}{2} \left(\frac{\lambda}{1 - \lambda} \right)^{1/2} (z^\alpha - Z) I(Z < z^\alpha),$$

2. The sequence $\{U_n\}$ is uniformly integrable, hence

$$\begin{aligned} ERR_x^* &= \lim_{n \rightarrow \infty} E [U_n] = E [U] \\ &= \frac{1}{2} \left(\frac{\lambda}{1 - \lambda} \right)^{1/2} \left[z^\alpha (1 - \alpha) + \frac{\exp \{ -(z^\alpha)^2 / 2 \}}{(2\pi)^{1/2}} \right]. \end{aligned}$$

Proof: To prove (1) first write

$$Z_n = (Var [M])^{-1/2} (M - n_x/2),$$

$$U_n = \frac{(c_n - M)I(M < c_n)}{(c'_n - M)I(M < c'_n)} \left(\frac{\text{Var}[M]}{n} \right)^{1/2} \left(1 - \frac{M}{n/2 + 1} \right)^{-1} (z^\alpha - Z_n) I(Z_n < z^\alpha),$$

where c'_n is the normal approximation to the critical value. As $n \rightarrow \infty$,

$$\frac{(c_n - M)I(M < c_n)}{(c'_n - M)I(M < c'_n)} \xrightarrow{p} 1,$$

$$\left(\frac{\text{Var}[M]}{n} \right)^{1/2} \rightarrow \frac{[\lambda(1-\lambda)]^{1/2}}{2}, \quad \left(1 - \frac{M}{n/2 + 1} \right)^{-1} \xrightarrow{p} (1-\lambda)^{-1},$$

the last convergence resulting from $M/n \xrightarrow{p} \lambda/2$. Z_n converges in law to the standard normal Z ; $(z^\alpha - x)I(x < z^\alpha)$ is an everywhere continuous function, as a result $(z^\alpha - Z_n)I(Z_n < z^\alpha)$ converges in law to a random variable having the distribution of $(z^\alpha - Z)I(Z < z^\alpha)$. An application of Slutsky's theorem proves part (1) of the theorem. (In Section 5.5 it will be shown directly that $n^{-1/2}R_n^x \xrightarrow{\mathcal{L}} U$.)

To verify the uniform integrability of U_n requires demonstrating, for arbitrary $\varepsilon > 0$, the existence of a constant b_\circ such that for $b > b_\circ$

$$\sup_n E[|U_n| I(|U_n| > b)] < \varepsilon.$$

Write

$$U_n = \left(\frac{\text{Var}[M]}{n} \right)^{1/2} \left(1 - \frac{M}{n/2 + 1} \right)^{-1} \left(\frac{c_n - M}{(\text{Var}[M])^{1/2}} \right) I(M < c_n).$$

The term $(\text{Var}[M]/n)^{1/2}$ is bounded above by 1. As $|U_n| > 0$ only if $M < c_n$,

$$\left(1 - \frac{M}{n/2 + 1} \right)^{-1} I(M < c_n) \leq \left(1 - \frac{c_n}{n/2 + 1} \right)^{-1}. \quad (5.3)$$

As $n \rightarrow \infty$ the right-hand side of (5.3) has a limiting value of $(1-\lambda)^{-1}$; take γ_1 as

an upper bound, also bounding the left-hand side. Then

$$|U_n| \leq \frac{\gamma_1 |c_n - M|}{(\text{Var}[M])^{1/2}}.$$

Further, $I(|U_n| > b)$ is bounded above by

$$I\left(\frac{\gamma_1 |c_n - M|}{(\text{Var}[M])^{1/2}} > b\right).$$

Let γ_2 be an upper bound for $(c_n - M)^2 / \text{Var}[M] \rightarrow (z^\alpha)^2$ and set $b_o = \gamma_1^2(\gamma_2 + 1)/\varepsilon$.

Then applying in succession the results above, Holder's Inequality and Markov's Inequality, for all $b > b_o$:

$$\begin{aligned} E[|U_n| I(|U_n| > b)] &\leq E\left[\left(\frac{\gamma_1 |c_n - M|}{(\text{Var}[M])^{1/2}}\right) I\left(\frac{\gamma_1 |c_n - M|}{(\text{Var}[M])^{1/2}} > b\right)\right] \\ &\leq \left\{E\left[\frac{\gamma_1^2 (c_n - M)^2}{\text{Var}[M]}\right]\right\}^{1/2} \left\{P\left[\frac{(c_n - M)^2}{\text{Var}[M]} > \frac{b^2}{\gamma_1^2}\right]\right\}^{1/2} \\ &\leq [\gamma_1^2(\gamma_2 + 1)]^{1/2} \left[\frac{\gamma_1^2(\gamma_2 + 1)}{b^2}\right]^{1/2} \\ &= \frac{\gamma_1^2(\gamma_2 + 1)}{b} \leq \frac{\gamma_1^2(\gamma_2 + 1)}{b_o} = \varepsilon. \end{aligned}$$

The lemma is proved. ■

Table 5.2 provides exact and approximate values for three sample sizes and three configurations of (n_x, n_y) . Due to the discrete nature of the statistic M , the level of each test was selected to be conservative relative to a nominal $\alpha = .05$. The exact values represent upper bounds for the expected minimal contamination.

Table 5.2: Expected Resistance to Rejection of Mood's Test: Smaller Sample Contamination, Exact and Approximate Values

(n_x, n_y)	α	<u>Exact</u>		<u>Approximate</u>	
		Count	Proportion	Count	Proportion
(15,15)	.0134	6.3181	.2106	6.0760	.2025
(30,30)	.0349	7.4229	.1237	7.0749	.1179
(60,60)	.0500	9.4546	.0788	9.1246	.0760
(10,20)	.0251	4.1967	.1399	3.8109	.1270
(20,40)	.0269	5.7223	.0954	5.3073	.0885
(40,80)	.0404	7.2655	.0605	6.8273	.0569
(5,25)	.0211	2.8846	.0962	2.4983	.0833
(10,50)	.0399	3.4993	.0583	3.0629	.0510
(20,100)	.0423	4.7151	.0393	4.2660	.0355

5.3 Optimal contamination scheme

The all X_i contamination has been shown to depend on the location of the $(c_n - M)$ th Y_j above $\hat{\eta}$. Similarly, the all Y_j contamination depends on the location of the $(c_n - M)$ th X_i below $\hat{\eta}$. The position of the medians after these contaminations will be denoted η_x^* and η_y^* respectively:

$$\eta_x^* = \inf \left\{ z : n_y \left[\hat{F}_y(z) - \hat{F}_y(\hat{\eta}) \right] = (c_n - M) I(M < c_n) \right\},$$

$$\eta_y^* = \inf \left\{ z : n_x \left[\hat{F}_x(z) - \hat{F}_x(\hat{\eta}) \right] = (M - c_n) I(M < c_n) \right\}.$$

Then $\eta_x^* = \eta_y^* = \hat{\eta}$ in case $M \geq c_n$.

Proposition 5.1 *Let R_n denote the optimal contamination. Define*

$$W_n(z) = 2(c_n - M) I(M < c_n) + n_x \left[\hat{F}_x(z) - \hat{F}_x(\hat{\eta}) \right] - n_y \left[\hat{F}_y(z) - \hat{F}_y(\hat{\eta}) \right].$$

Then

$$R_n = \min_{\eta_y^* \leq z \leq \eta_x^*} W_n(z).$$

The position of the median after the optimal contamination, denoted η° , is then, for uniqueness, given by the infimum of all z such that $W_n(z) = R_n$; $\eta^\circ = \hat{\eta}$ when $M \geq c_n$, and $\eta_y^ \leq \eta^\circ \leq \eta_x^*$.*

Moving from left to right the function $W_n(z)$ increases by 1 on each X_i and decreases by 1 on each Y_j . Note that $W_n(\eta_x^*) = R_n^x$, $W_n(\eta_y^*) = R_n^y$, and that $M \geq c_n$ implies $R_n = 0$.

Let c be an integer. Consider any contamination scheme of a X_i 's and b Y_j 's where $(a - b) = c, a \geq 0, b \geq 0$. Subject to this constraint we minimize $(a + b)$.

The first observation is that an optimal contamination cannot occur with $c > R_n^x$. If $c > R_n^x$ then $a > R_n^x$, implying $(a + b) > R_n^x$. Similarly, $c < -R_n^y$ cannot lead to an optimal solution.

First assume $c > 0$. We can minimize $(a + b)$ subject to $(a - b) = c$ by minimizing a . With $(a - b) = c$ fixed the median after the contamination will be shifted c data items right. Let x be the number of these items that are X_i ; $y = (c - x)$ are then Y_j . In order to force rejection we must have $(M + a - x) \geq c_n$. The minimal value of a is given by $(c_n - M + x)$, with a total contamination of $2(c_n - M) + (x - y)$. The same is true for $c \leq 0$ (the case $c = 0$ represents the strategy of “switching” $(c_n - M)$ X_i ’s below $\hat{\eta}$ with $(c_n - M)$ Y_j ’s below $\hat{\eta}$).

Take $\eta(c)$ to be the c th observation above (below, if $c < 0$) $\hat{\eta}$. Then $2(c_n - M)I(M < c_n) + (x - y) = W_n(\eta(c))$. The problem then reduces to

$$R_n = \min_{-R_n^y \leq c \leq R_n^x} W_n(\eta(c)),$$

which is equivalent to the statement of the proposition as $W_n(\cdot)$ is constant on the intervals $[\eta(c), \eta(c + 1))$, $\eta(R_n^x) = \eta_x^*$, and $\eta(R_n^y) = \eta_y^*$. ■

5.4 Bounding results

This section is devoted to bounding random variables that play a part in determining various contamination schemes. The development to follow will make great use of the “after contamination medians” η_x^* and η_y^* . The general attack begins by placing almost sure bounds on $(\hat{\eta} - \eta)$ and $(M - n_x/2)$; this will imply bounds on $(\eta_x^* - \eta)$ and $(\eta_y^* - \eta)$. As $\eta_y^* \leq \eta^\circ \leq \eta_x^*$, with probability one we will be able to place each of these after contamination medians in a sufficiently small neighborhood of η .

Partition that neighborhood, and show that, with probability one, each segment of the partition contains essentially the expected number of both X_i and Y_j as $n \rightarrow \infty$. For the case $\lambda < 1/2$ the conclusion is that, for any $\varepsilon \in (0, 1/2)$, with probability one, $(R_n - R_n^x) = O(n^\varepsilon), n \rightarrow \infty$. For the case $\lambda = 1/2$ the result is also true provided $\varepsilon \in (1/4, 1/2)$.

Begin by placing almost sure bounds on the Mood statistic. Hoeffding's Inequality is needed.

Lemma 5.2 (Hoeffding) *Let B be a binomial random variable, m trials and p probability. Then*

$$P[(B - E\{B\}) \geq mt] \leq \exp\{-2mt^2\}.$$

Lemma 5.3 *There exists a constant K_1 such that*

$$P\left[\left|M - \frac{n_x}{2}\right| \geq K_1 n^{1/2} (\log n)^{1/2} \text{ i.o.}\right] = 0. \quad (5.4)$$

Proof: Recall from (1.8) that

$$\left(M - \frac{n_x}{2}\right) = \frac{n_x n_y}{n} \left\{ [\hat{F}_x(\eta) - 1/2] - [\hat{F}_y(\eta) - 1/2] + R_n \right\},$$

where, with probability one, $R_n = O(n^{-3/4} (\log n)^{3/4})$ as $n \rightarrow \infty$. Using Lemma 5.2,

$$\begin{aligned} & P\left[|\hat{F}_x(\eta) - 1/2| \geq n_x^{-1/2} (\log n_x)^{1/2}\right] \\ &= 2P\left[n_x \left(\hat{F}_x(\eta) - 1/2\right) \geq n_x n_x^{-1/2} (\log n_x)^{1/2}\right] \\ &\leq 2 \exp\left\{-2n_x \left[n_x^{-1/2} (\log n_x)^{1/2}\right]^2\right\} = 2n_x^{-2}. \end{aligned}$$

As

$$\sum_{n_x=1}^{\infty} P \left[\left| \hat{F}_x(\eta) - 1/2 \right| \geq n_x^{-1/2} (\log n_x)^{1/2} \right] \leq \sum_{n_x=1}^{\infty} \frac{2}{n_x^2} < \infty,$$

by the Borel-Cantelli Lemma,

$$P \left[\left| \hat{F}_x(\eta) - 1/2 \right| \geq n_x^{-1/2} (\log n_x)^{1/2} \text{ i.o.} \right] = 0.$$

An identical argument establishes an analogous result for $|\hat{F}_y(\eta) - 1/2|$. Since, with probability one, $R_n = o(n^{-1/2}(\log n)^{1/2})$, $n \rightarrow \infty$, bounding it is trivial. There then exist constants K_2 , K_3 , and K_4 such that

- $P \left[n \left| \hat{F}_x(\eta) - 1/2 \right| \geq K_2 n^{1/2} (\log n)^{1/2} \text{ i.o.} \right] = 0,$
- $P \left[n \left| \hat{F}_y(\eta) - 1/2 \right| \geq K_3 n^{1/2} (\log n)^{1/2} \text{ i.o.} \right] = 0,$ and
- $P \left[n |R_n| \geq K_4 n^{1/2} (\log n)^{1/2} \text{ i.o.} \right] = 0.$

Take $K_1 = K_2 + K_3 + K_4$. Then

$$\begin{aligned} & \left[\omega : \left| M - \frac{n_x}{2} \right| \geq K_1 n^{1/2} (\log n)^{1/2} \right] \\ &= \left[\omega : \frac{n_x n_y}{n} \left\{ \left| \left[\hat{F}_x(\eta) - 1/2 \right] - \left[\hat{F}_y(\eta) - 1/2 \right] + R_n \right| \right\} \geq K_1 n^{1/2} (\log n)^{1/2} \right] \\ &\subset \left[\omega : n \left\{ \left| \hat{F}_x(\eta) - 1/2 \right| + \left| \hat{F}_y(\eta) - 1/2 \right| + |R_n| \right\} \geq K_1 n^{1/2} (\log n)^{1/2} \right] \\ &\subset \left\{ \left[\omega : n \left| \hat{F}_x(\eta) - 1/2 \right| \geq K_2 n^{1/2} (\log n)^{1/2} \right] \right. \\ &\quad \cup \left[\omega : n \left| \hat{F}_y(\eta) - 1/2 \right| \geq K_3 n^{1/2} (\log n)^{1/2} \right] \\ &\quad \left. \cup \left[\omega : n |R_n| \geq K_4 n^{1/2} (\log n)^{1/2} \right] \right\}. \end{aligned}$$

Taking the limit supremum on both sides, then probabilities, proves the result. (The law of the iterated logarithm can be implemented in place of Hoeffding's Inequality, thus ensuring a slightly tighter result.) ■

As a consequence of this lemma, noting that $(c_n - n_x/2) = O(n^{1/2})$, there exists a constant K_5 such that

$$P \left[\frac{c_n - M}{n_y} \leq K_5 n^{-1/2} (\log n)^{1/2} \text{ a.a.} \right] = 1. \quad (5.5)$$

For reference purposes Lemmas 1.1 and 1.2 are restated.

- There exists a constant K_6 such that

$$P \left[|\hat{\eta} - \eta| \leq K_6 n^{-1/2} (\log n)^{1/2} \text{ a.a.} \right] = 0. \quad (5.6)$$

- Let $\{a_{n_y}\}$ be a sequence of constants, $a_{n_y} \sim C_0 n_y^{1/2} (\log n_y)^{1/2}$, where $C_0 > 0$.

Then there exists a constant K_7 (depending on C_0) such that

$$P \left\{ \sup_{|x| \leq a_{n_y}} \left| \left[\hat{F}_y(\eta + x) - \hat{F}_y(\eta) \right] - x \right| \leq K_7 n^{-3/4} (\log n)^{3/4} \text{ a.a.} \right\} = 0. \quad (5.7)$$

This result can also be stated in terms of X_i .

The results (5.5), (5.6) and (5.7) are sufficient to place an almost sure bound on $(\eta_x^* - \hat{\eta})$.

Lemma 5.4 *There exists a constant K and set Ω_0 , $P[\Omega_0] = 1$ such that for all $\omega \in \Omega_0$*

$$(\eta_x^* - \hat{\eta}) \leq K n^{-1/2} (\log n)^{1/2}$$

for n sufficiently large.

Proof: By definition $(\eta_x^* - \hat{\eta}) \geq 0$. If $M \geq c_n$ then $(\eta_x^* - \hat{\eta}) = 0$. It suffices to consider only the case $M < c_n$.

Take $K > K_5$. Restrict analysis to the set Ω_0 , $P[\Omega_0] = 1$, defined as the intersection of the sets on which (by (5.5), (5.6) and (5.7)) we have, for n sufficiently large,

- $(c_n - M)/n_y \leq K_5 n^{-1/2} (\log n)^{1/2}$,
- $|\hat{\eta} - \eta| \leq K_6 n^{-1/2} (\log n)^{1/2}$, and
- $\sup_{x \leq a_{n_y}} |[\hat{F}_y(\eta + x) - \hat{F}_y(\eta)] - x| \leq K_7 n^{-3/4} (\log n)^{3/4}$, where $\{a_{n_y}\}$ is chosen with $a_{n_y} \geq (K + K_6)n^{-1/2} (\log n)^{1/2}$.

As a result, $(\hat{\eta} + K n^{-1/2} (\log n)^{1/2} - \eta) \in (-a_{n_y}, a_{n_y})$; then

$$\left\{ \left| \left[\hat{F}_y \left(\hat{\eta} + K n^{-1/2} (\log n)^{1/2} \right) - \hat{F}_y(\eta) \right] - \left[\hat{\eta} + K n^{-1/2} (\log n)^{1/2} - \eta \right] \right| \right\} \leq K_7 n^{-3/4} (\log n)^{3/4}. \quad (5.8)$$

Similarly,

$$\left| \left[\hat{F}_y(\hat{\eta}) - \hat{F}_y(\eta) \right] - [\hat{\eta} - \eta] \right| \leq K_7 n^{-3/4} (\log n)^{3/4}. \quad (5.9)$$

Combining (5.8) and (5.9),

$$\left[\hat{F}_y \left(\hat{\eta} + K n^{-1/2} (\log n)^{1/2} \right) - \hat{F}_y(\hat{\eta}) \right] \geq K n^{-1/2} (\log n)^{1/2} - 2K_7 n^{-3/4} (\log n)^{3/4}. \quad (5.10)$$

As $K > K_5$, it can also be assumed that n is large enough to force the right-hand

side of (5.10) to exceed $Kn^{-1/2}(\log n)^{1/2}$. Then

$$\left[\hat{F}_y \left(\hat{\eta} + Kn^{-1/2}(\log n)^{1/2} \right) - \hat{F}_y(\hat{\eta}) \right] \geq Kn^{-1/2}(\log n)^{1/2} \geq \frac{c_n - M}{n_y}. \quad (5.11)$$

Now, $(\eta_x^* - \hat{\eta}) \leq Kn^{-1/2}(\log n)^{1/2}$ if and only if there are at least $(c_n - M) Y_j$ in the interval $(\hat{\eta}, \hat{\eta} + Kn^{-1/2}(\log n)^{1/2})$. This is precisely what (5.11) assures for n sufficiently large—establishing the lemma. ■

Corollary 5.1 *There exists a constant K and set Ω_0 , $P[\Omega_0] = 1$, such that for all $\omega \in \Omega_0$,*

$$\max \left\{ |\eta_x^* - \eta|, |\eta_y^* - \eta| \right\} \leq Kn^{-1/2}(\log n)^{1/2} \quad (5.12)$$

for n sufficiently large.

Proof: $(\hat{\eta} - \eta_y^*)$ can be bounded as in Lemma 5.4. As $|\eta_x^* - \eta| \leq |\eta_x^* - \hat{\eta}| + |\hat{\eta} - \eta|$, the result follows immediately from (5.6) and Lemma 5.4. ■

Both η_x^* and η_y^* converge to η almost surely.

5.5 Limiting distribution of the single sample scheme

We show that $n^{-1/2}R_n^x$ can be written as function of $(c_n - M)$ plus negligible remainder.

Take K as in Corollary 5.1 and apply the uniform convergence result (5.7) on the X_i ($a_{n_x} > Kn^{-1/2}(\log n)^{1/2}$) and the Y_j ($a_{n_y} > Kn^{-1/2}(\log n)^{1/2}$). Appealing to the method of Theorem 1.1, with probability one,

$$\left[\hat{F}_x(\eta_x^*) - \hat{F}_x(\hat{\eta}) \right] - \left[\hat{F}_y(\eta_x^*) - \hat{F}_y(\hat{\eta}) \right] = O \left(n^{-3/4}(\log n)^{3/4} \right), n \rightarrow \infty. \quad (5.13)$$

As $(c_n - M)/n_y = [\hat{F}_y(\eta_x^*) - \hat{F}_y(\hat{\eta})]$, with probability one, R_n^x , the value of the all X_i contamination, the sum of $(c_n - M)$ and the number of $X_i \in (\hat{\eta}, \eta_x^*]$, can be expressed

$$\begin{aligned}
R_n^x &= (c_n - M) I(M < c_n) + n_x [\hat{F}_x(\eta_x^*) - \hat{F}_x(\hat{\eta})] \\
&= (c_n - M) I(M < c_n) + n_x \left\{ [\hat{F}_y(\eta_x^*) - \hat{F}_y(\hat{\eta})] + O\left(n^{-3/4} (\log n)^{3/4}\right) \right\} \\
&= \frac{n}{n_y} (c_n - M) I(M < c_n) + O\left(n^{1/4} (\log n)^{3/4}\right), n \rightarrow \infty. \tag{5.14}
\end{aligned}$$

Lemma 5.5 *Assume $n_x/n \rightarrow \lambda$. Then as $n \rightarrow \infty$, $n^{-1/2} R_n^x$ converges in law to a random variable having the distribution of U (as given in Lemma 5.1). Hence, the sequence $\{n^{-1/2} R_n^x\}$ is uniformly integrable.*

Proof: The representation of (5.14) and an application of Slutsky's theorem proves the first assertion. The second follows from a standard result, see Serfling (1980), page 15, Lemma B. ■

5.6 A partition

The next result partitions an $O(n^{-1/2} (\log n)^{1/2})$ neighborhood of η and places usable bounds on the excess of X_i over Y_j in each segment of the partition. Bernstein's inequality will be needed.

Lemma 5.6 (Bernstein) *Let B be a binomial random variable, m trials and p probability. Then*

$$P[|B - mp| \geq mt] \leq 2 \exp \left\{ -\frac{mt^2}{2(p+t)} \right\}.$$

For positive integers n define $a_n = An^{-1/2} (\log n)^{1/2}$ where $A > 0$. For fixed $\varepsilon \in (0, 1/2)$ define:

$$b_n = \lfloor An^{1/2-\varepsilon} (\log n)^{1/2} \rfloor,$$

where $\lfloor \cdot \rfloor$ denotes the “floor” function. For each n partition the interval $[-a_n, a_n]$ into $2b_n$ segments, each having length $n^{\varepsilon-1}$. For $k = -b_n, \dots, b_n$ the partition points are given by $\eta_n(k) = kn^{\varepsilon-1}$, with $\eta_n(0) = \eta$. There are then $2b_n$ segments $(\eta_n(k-1), \eta_n(k))$ partitioning $[-a_n, a_n]$.

For each $k \in \{-b_n + 1, \dots, b_n\}$ set

$$D_n(k) = \left| n_x \left[\hat{F}_x(\eta_n(k)) - \hat{F}_x(\eta_n(k-1)) - n^{\varepsilon-1} \right] - n_y \left[\hat{F}_y(\eta_n(k)) - \hat{F}_y(\eta_n(k-1)) - n^{\varepsilon-1} \right] \right|,$$

and put

$$Z_n = \max_{k \in \{-b_n+1, \dots, b_n\}} \{D_n(k)\}.$$

Lemma 5.7 *With probability one, $Z_n = O(n^{\varepsilon/2} (\log n)^{1/2})$ as $n \rightarrow \infty$.*

Proof: By the Borel-Cantelli Lemma, it is sufficient to exhibit values L and n_0 such that

$$\sum_{n=n_0}^{\infty} P \left[Z_n \geq Ln^{\varepsilon/2} (\log n)^{1/2} \right] < \infty.$$

These values will be provided during the course of the proof.

Begin as follows:

$$P \left[Z_n \geq Ln^{\varepsilon/2} (\log n)^{1/2} \right]$$

$$\begin{aligned}
&= P \left[\max_{k \in \{-b_n+1, \dots, b_n\}} \{D_n(k)\} \geq Ln^{\varepsilon/2} (\log n)^{1/2} \right] \\
&\leq \sum_{k=-b_n+1}^{b_n} P \left[D_n(k) \geq Ln^{\varepsilon/2} (\log n)^{1/2} \right] \\
&= 2b_n P \left[D_n(1) \geq Ln^{\varepsilon/2} (\log n)^{1/2} \right] \\
&\leq 2An^{1/2-\varepsilon} (\log n)^{1/2} P \left\{ \left| n_x \left[\hat{F}_x(\eta_n(1)) - \hat{F}_x(\eta_n(0)) - n^{\varepsilon-1} \right] \right. \right. \\
&\quad \left. \left. - n_y \left[\hat{F}_y(\eta_n(1)) - \hat{F}_y(\eta_n(0)) - n^{\varepsilon-1} \right] \right| \geq Ln^{\varepsilon/2} (\log n)^{1/2} \right\} \\
&\leq 2An^{1/2-\varepsilon} (\log n)^{1/2} \times \\
&\quad \left\{ P \left[n_x \left| \hat{F}_x(\eta_n(1)) - \hat{F}_x(\eta_n(0)) - n^{\varepsilon-1} \right| \geq Ln^{\varepsilon/2} (\log n)^{1/2} / 2 \right] \right. \\
&\quad \left. + P \left[n_y \left| \hat{F}_y(\eta_n(1)) - \hat{F}_y(\eta_n(0)) - n^{\varepsilon-1} \right| \geq Ln^{\varepsilon/2} (\log n)^{1/2} / 2 \right] \right\}. \quad (5.15)
\end{aligned}$$

Apply Lemma 5.6 with $p = n^{\varepsilon-1}$ and $t = Ln^{\varepsilon/2} (\log n)^{1/2} / 2n_x$. Then

$$P \left[n_x \left| \hat{F}_x(\eta_n(1)) - \hat{F}_x(\eta_n(0)) - n^{\varepsilon-1} \right| \geq Ln^{\varepsilon/2} (\log n)^{1/2} / 2 \right] \leq 2 \exp \{-\tau_n\},$$

where

$$\tau_n = \frac{L^2 \log n}{8n_x/n + 4Ln^{-\varepsilon/2} (\log n)^{1/2}} \geq \frac{L^2 \log n}{8 + 4Ln^{-\varepsilon/2} (\log n)^{1/2}}.$$

Let n_0 be the value at which $n^{-\varepsilon/2} (\log n)^{1/2}$ attains its maximum, let C be that maximum. Choose L such that $L^2/(8 + 4LC) = 2$. Then, for $n \geq n_0$, $\tau_n \geq 2 \log n$, and

$$P \left[n_x \left| \hat{F}_x(\eta_n(1)) - \hat{F}_x(\eta_n(0)) - n^{\varepsilon-1} \right| \geq Ln^{\varepsilon/2} (\log n)^{1/2} / 2 \right] \leq 2n^{-2}.$$

An identical statement holds for the probability in (5.15). Therefore

$$\sum_{n=n_0}^{\infty} P \left[Z_n \geq Ln^{\varepsilon/2} (\log n)^{1/2} \right] \leq \sum_{n=n_0}^{\infty} 8An^{-3/2-\varepsilon} (\log n)^{1/2} < \infty,$$

completing the proof. ■

Define

$$D_n^x(k) = \left| n_x \left[\hat{F}_x(\eta_n(k)) - \hat{F}_x(\eta_n(k-1)) - n^{\varepsilon-1} \right] \right|,$$

$$Z_n^x = \max_{k \in \{-b_n+1, \dots, b_n\}} \{D_n^x(k)\}.$$

A proof similar to that given in Lemma 5.7 exhibits a constant L_2 such that

$$P \left[Z_n^x \geq L_2 n^{\varepsilon/2} (\log n)^{1/2} \text{ i.o.} \right] = 0. \quad (5.16)$$

5.7 Equivalency of the optimal and smaller sample schemes

The partitioning result of the previous section will be used to show that, in the case $\lambda < 1/2$, for any $\varepsilon \in (0, 1/2)$, with probability one, the optimal scheme differs from the all X_i scheme by $O(n^\varepsilon)$, $n \rightarrow \infty$. In the case $\lambda = 1/2$ the bound is not as tight; $\varepsilon > 1/4$ suffices.

From this point on the index k shall be used to refer to the k th segment of the partition, $(\eta_n(k-1), \eta_n(k)]$, where k is understood to satisfy $-b_n + 1 \leq k \leq b_n$.

Take $A = K + 1$ where K is defined as in Corollary 5.1. For all $n \geq 2$ we have $\eta_n(-b_n) \leq -Kn^{-1/2} (\log n)^{1/2} \leq Kn^{-1/2} (\log n)^{1/2} \leq \eta_n(b_n)$. The segments of the partition cover the entire neighborhood of values in which, with probability one, η_y^* and η_x^* lie for n sufficiently large.

Define $\Delta_n(k)$ to be the excess of X_i over Y_j in segment k :

$$\Delta_n(k) = n_x [\hat{F}_x(\eta_n(k)) - \hat{F}_x(\eta_n(k-1))] - n_y [\hat{F}_y(\eta_n(k)) - \hat{F}_y(\eta_n(k-1))].$$

For a value z contained in the k th segment of the partition define

$$\Delta_n^+(z, k) = n_x [\hat{F}_x(\eta_n(k)) - \hat{F}_x(z)] - n_y [\hat{F}_y(\eta_n(k)) - \hat{F}_y(z)],$$

$$\Delta_n^-(z, k) = n_x [\hat{F}_x(z) - \hat{F}_x(\eta_n(k-1))] - n_y [\hat{F}_y(z) - \hat{F}_y(\eta_n(k-1))].$$

$\Delta_n^+(z, k)$ and $\Delta_n^-(z, k)$ count the excess of X_i over Y_j in partition k and, respectively, to the right and left of z .

From Proposition 5.1 we have $R_n = W_n(\eta^\circ)$, and $R_n^x = W_n(\eta_x^*)$, then

$$0 \leq R_n^x - R_n = n_x [\hat{F}_x(\eta_x^*) - \hat{F}_x(\eta^\circ)] - n_y [\hat{F}_y(\eta_x^*) - \hat{F}_y(\eta^\circ)]. \quad (5.17)$$

5.7.1 Unequal sample sizes

Lemma 5.8 *Suppose $\lambda < 1/2$. Then for any $\varepsilon \in (0, 1/2)$, with probability one $R_n^x - R_n = O(n^\varepsilon)$, as $n \rightarrow \infty$.*

Proof: Take Ω_0 to be the set on which, by Corollary 5.1, Lemma 5.7, and (5.16), for n sufficiently large,

1. $\max\{|\eta_x^* - \eta|, |\eta_y^* - \eta|\} \leq Kn^{-1/2}(\log n)^{1/2}$,
2. $Z_n \leq Ln^{\varepsilon/2}(\log n)^{1/2}$, with $A > K + 1$, and
3. $Z_n^x \leq L_2 n^{\varepsilon/2}(\log n)^{1/2}$, with $A > K + 1$.

Suppose that η° and η_x^* lie in segments k_1 and k_2 of the partition, respectively

($k_1 \leq k_2$). Split $R_n^x - R_n$, as in (5.17), over the partition,

$$0 \leq R_n^x - R_n = \Delta_n^+(\eta^\circ, k_1) + \sum_{k=k_1+1}^{k_2-1} \Delta_n(k) + \Delta_n^-(\eta_x^*, k_2).$$

Using the monotonicity of \hat{F}_x we have

$$\begin{aligned} \Delta_n^+(\eta^\circ, k_1) &\leq n_x [\hat{F}_x(\eta_n(k+1)) - \hat{F}_x(\eta_n(k))] \\ &\leq \left(\frac{n_x}{n}\right) n^\varepsilon + L_2 n^{\varepsilon/2} (\log n)^{1/2}. \end{aligned}$$

The same result holds for $\Delta_n^-(\eta_x^*, k_2)$. We also have, for all k ,

$$\Delta_n(k) \leq \left(\frac{n_x - n_y}{n}\right) n^\varepsilon + L n^{\varepsilon/2} (\log n)^{1/2}.$$

Hence, with probability one, for n sufficiently large

$$0 \leq R_n^x - R_n \leq n^\varepsilon \left\{ O(1) + (k_2 - k_1 + 1) \left[\frac{n_x - n_y}{n} + L n^{-\varepsilon/2} (\log n)^{1/2} \right] \right\}.$$

As $n \rightarrow \infty$, $(n_x - n_y)/n \rightarrow 2\lambda - 1 < 0$ and $n^{-\varepsilon/2} (\log n)^{1/2} \rightarrow 0$. As a result,

$$0 \leq R_n^x - R_n \leq n^\varepsilon O(1),$$

implying the statement of the lemma. ■

5.7.2 Equal sample sizes

Take $\lambda = 1/2$. Applying the uniform convergence result (5.7) again, with probability one,

$$0 \leq R_n^x - R_n = n_x [\hat{F}_x(\eta_x^*) - \hat{F}_x(\eta^\circ)] - n_y [\hat{F}_y(\eta_x^*) - \hat{F}_y(\eta^\circ)]$$

$$\begin{aligned}
&= n_x \left\{ \left[\hat{F}_x(\eta_x^*) - \hat{F}_x(\eta^\circ) \right] - \left[\hat{F}_y(\eta_x^*) - \hat{F}_y(\eta^\circ) \right] \right\} \\
&\quad + (n_y - n_x) \left[\hat{F}_y(\eta_x^*) - \hat{F}_y(\eta^\circ) \right] \\
&= n_x O\left(n^{-3/4}(\log n)^{3/4}\right) + O(1) = O\left(n^{1/4}(\log n)^{3/4}\right), n \rightarrow \infty,
\end{aligned}$$

as $n_y - n_x = O(1)$ has been assumed. Lemma 5.8 can be extended to all cases by taking $\varepsilon > 1/4$.

5.7.3 Equivalence

Theorem 5.1 $n^{-1/2}(R_n^x - R_n) \rightarrow 0$ almost surely. Hence the limiting distribution of $n^{-1/2}R_n$ is that of the random variable U and

$$ERR^* = \lim_{n \rightarrow \infty} E \left[n^{-1/2} R_n \right] = ERR_x^*.$$

Proof: Lemmas 5.5 and 5.8 are sufficient to prove that $n^{-1/2}R_n \xrightarrow{\mathcal{L}} U$. The random variable $n^{-1/2}(R_n^x - R_n)$ is bounded above by $n^{-1/2}R_n^x$. Lemma 5.5 demonstrated the uniform integrability of $\{n^{-1/2}R_n^x\}$, hence $\{n^{-1/2}(R_n^x - R_n)\}$ is uniformly integrable. As $n^{-1/2}(R_n^x - R_n) \xrightarrow{\mathcal{L}} 0$, it follows by a standard result (see Serfling (1980), page 14, Theorem A) that $\lim_{n \rightarrow \infty} E \left[n^{-1/2}(R_n^x - R_n) \right] = 0$, implying the statement of the theorem. ■

5.8 Distance to rejection

In light of the results of the previous section we now examine a property of the small sample contamination scheme. Assume that $n_x/n \rightarrow \lambda \leq 1/2$. We have, with probability one, $(\eta_x^* - \eta_y^*) = O(n^{-1/2}(\log n)^{1/2})$ as $n \rightarrow \infty$. An immediate

implication of (5.14) is that, with probability one, $R_n^x = O(n^{1/2} (\log n)^{1/2})$ as $n \rightarrow \infty$.

Consider then the strategy of forcing rejection by placing all X_i between η_y^* and η_x^* , inclusive, at η_x^* —denote this strategy by \mathcal{S} . Let \mathcal{I} denote the index set of the altered X_i . Define the *distance to rejection* of the strategy \mathcal{S} by $d_n(\mathcal{S}) = \sum_{\mathcal{I}} (\eta_x^* - X_i)$. Then, with probability one,

$$\begin{aligned} d_n(\mathcal{S}) &= \sum_{\mathcal{I}} (\eta_x^* - X_i) \leq R_n^x (\eta_x^* - \eta_y^*) \\ &= O(n^{1/2} (\log n)^{1/2}) O(n^{-1/2} (\log n)^{1/2}) = O(\log n), n \rightarrow \infty. \end{aligned}$$

Additionally, $d_n(\mathcal{S}) = O_p(1)$ as $n \rightarrow \infty$.

The result is true for any F twice differentiable in a neighborhood of the median. As $n \rightarrow \infty$ the neighborhood of η under consideration can be made suitably uniform. For any strongly consistent estimate, $\tilde{\eta}$, of η , with probability one,

$$F(\tilde{\eta}) - F(\eta) = f(\eta)(\tilde{\eta} - \eta) + O((\tilde{\eta} - \eta)^2), n \rightarrow \infty.$$

In particular $O((\eta_x^* - \eta)^2) = O((\eta_y^* - \eta)^2) = O(n^{-1} \log n)$.

The result contrasts nicely with the test based on the sample mean. Adopting the strategy, \mathcal{T} , that breaks down the t -test in no more than a single contamination, a straightforward application of the law of the iterated logarithm gives the distance to rejection for the normal theory test as satisfying, with probability one, $d_n(\mathcal{T}) = O((n \log \log n)^{1/2})$, as $n \rightarrow \infty$. Further, adopting the contamination scheme \mathcal{S} , the limiting distribution of the normal theory test statistic is unaltered.

With distance to rejection as the criteria, sign-based testing methods are clearly inferior.

5.9 Remark

A similar result can be derived for the 2-sided testing problem. In particular,

$$\text{ERR}^* = \frac{1}{2} \left(\frac{\lambda}{1-\lambda} \right)^{1/2} \left[z^{\alpha/2} (1-\alpha) - \sqrt{\frac{2}{\pi}} \left(1 - \exp \left\{ - (z^{\alpha/2})^2 / 2 \right\} \right) \right].$$

The two-sided test is less resistant to rejection than is the one-sided test.

Chapter 6

Resistance To Rejection of the Wilcoxon Signed Rank Test

The Wilcoxon signed rank test is based on the statistic

$$\sum_{i=1}^{n-1} \sum_{j=i}^n I\left(\frac{1}{2}(X_i + X_j) > \eta_0\right)$$

where X_1, \dots, X_n is a sample from a symmetric distribution with center of symmetry η . The test is used to test the hypothesis $\eta = \eta_0$ against shift alternatives. The asymptotic analysis of this procedure is more complicated than the analysis of the one-sample sign test, because the present procedure is based on ordered dependent random variables, namely the ordered $(X_i + X_j)/2$ for $i \leq j \leq n$.

Fortunately the theory of U -statistics can be applied. The statistic

$$V_n = \binom{n}{2}^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^n I\left(\frac{1}{2}(X_i + X_j) > \eta_0\right) \quad (6.1)$$

is a one-sample U -statistic and the test based on it is asymptotically equivalent to the Wilcoxon one-sample test. For these reasons, the test based on (6.1) will be considered first.

We assume that X_1, X_2, \dots is a sequence of independent random variables with common cumulative distribution function $F(x - \eta)$ where F is symmetric about 0. F has density f satisfying $\int f^2(x) dx < \infty$. $G(x - \eta)$ denotes the cumulative distribution function of $(X_1 + X_2)/2$, assume G twice differentiable on $(-\infty, +\infty)$.

6.1 Preliminaries

The following two results are needed.

Lemma 6.1 (Hoeffding (1963)) *Denote by U_n the statistic*

$$\binom{n}{r}^{-1} \sum_{\mathcal{C}} \psi(X_{i_1}, \dots, X_{i_r})$$

where X_1, X_2, \dots, X_n are i.i.d. random variables, ψ is a function symmetric in its r variables and \mathcal{C} denotes summation over all “ n choose r ” combinations of r X_i ’s.

Then

$$P(|U_n - E[U_n]| \geq t) \leq 2e^{-h} \quad (6.2)$$

where

$$h = \frac{kt^2}{2\left(\sigma^2 + \frac{1}{3}t \max(z, 1-z)\right)}$$

if $|\psi(x_1, \dots, x_r)| \leq 1, t \geq 0, k = \lfloor n/r \rfloor$, with

$$z = E\psi(X_1, \dots, X_r), \text{ and } \sigma^2 = \text{Var}\psi(X_1, \dots, X_r).$$

Lemma 6.1 is an extension of the Bernstein inequality used in Chapter 5. Next is the law of the iterated logarithm for U statistics, stated in Serfling (1980), page 191.

Lemma 6.2 *With probability one,*

$$\overline{\lim}_{n \rightarrow \infty} \frac{n^{1/2}(U_n - 1/2)}{(2r^2 \zeta_1 \log \log n)^{1/2}} = 1,$$

where $\zeta_1 = \text{Var}_F[\psi_1(X_1)] > 0$, with $\psi_1(x_1) = E_F[\psi(x_1, X_2)]$.

Applying the probability integral transformation, $W_i = F(X_i)$, the distributional properties of V_n remain unchanged; the W_i are now independent random variables, uniformly distributed over $[0, 1]$, and $\eta_0 = 1/2$. Again we are testing the one-sided alternative, $H_1 : \eta > \eta_0 = 1/2$; c_n is a sequence of critical values with $(3n)^{1/2}(c_n - 1/2) \rightarrow z^\alpha$.

6.2 Contamination points

To break down the Wilcoxon statistic, forcing rejection in every case, the lower-most sample values will be altered to values of, say, $+\infty$ (1 would suffice in this setting). Suppose we choose a value $0 \leq x \leq 1/2$ and force each sample value no greater than x to $+\infty$. In doing so, the net change in the value of V_n is given by

$$L_n(x) = \binom{n}{2}^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^n I[X_i + X_j \leq 1, (X_i \leq x \text{ or } X_j \leq x)].$$

To see this, observe that any Walsh average, $(X_i + X_j)/2$, $i < j$, involving an observation less than x will be forced above $1/2$ after the contamination. However, any pair (X_i, X_j) with Walsh average greater than $1/2$ prior to contamination will not be altered by the contamination. (The definition of L_n can be extended to $[0, 1]$ by taking $L_n(x) = L_n(1/2)$ for $x > 1/2$.) Note that $L_n(x)$ is a U statistic with symmetric kernel $\psi(x_1, x_2) = I(x_1 + x_2 \leq 1, (x_1 < x \text{ or } x_2 < x))$. Take $L(x) = EL_n(x) = E\psi(X_1, X_2) = 2x(1 - x)$. Define the *contamination point* by

$$X^* = \begin{cases} \inf \{x : L_n(x) \geq c_n - V_n\} & V_n < c_n, \\ 0 & \text{otherwise.} \end{cases}$$

X^* is the smallest x for which the test statistic is forced to rejection. Then R_n , the minimum contamination, is given by $nF_n(X^*)$.

6.3 Limiting distribution

The methods of the previous chapter are followed. In (5.14) it is shown that the necessary contamination, R_n^x , can be written as a function of the test statistic. The same result holds for this situation.

The following uniform convergence result makes use of techniques from quantile representation theory.

Lemma 6.3 *Let $\{e_n\}$ be a sequence of constants with $e_n \sim C_0 n^{-1/2} \log n$ as $n \rightarrow \infty$, for some fixed $C_0 > 0$. Let $B_n(x) = (L_n(x) - L(x))$, $I_n = (0, e_n)$ and $H_n = \sup_{x \in I_n} \{|B_n(x)|\}$. Then, with probability one, $H_n = O(n^{-3/4} \log n)$ as $n \rightarrow \infty$.*

Proof: Take $\{d_n\}$ to be a sequence of positive integers with $d_n \sim n^{1/4}$. Set $\eta_n(r) = r e_n / d_n$ where r is an integer, $0 \leq r \leq \lfloor d_n \rfloor$,

$$J_n(r) = [\eta_n(r), \eta_n(r+1)], \quad \alpha_n(r) = L(\eta_n(r+1)) - L(\eta_n(r)).$$

Then, for $x \in J_n(r)$, since both L_n and L are nondecreasing,

$$B_n(x) \leq L_n(\eta_n(r+1)) - L(\eta_n(r)) = B_n(\eta_n(r+1)) + \alpha_n(r).$$

Similarly $B_n(x) \geq B_n(\eta_n(r)) - \alpha_n(r)$ for $x \in J_n(r)$. Set

$$K_n = \max_{0 \leq r \leq d_n} \{|B_n(\eta_n(r))|\}, \quad \beta_n = \max_{0 \leq r \leq d_n-1} \{\alpha_n(r)\}.$$

Then $H_n \leq K_n + \beta_n$. Now,

$$\begin{aligned}\beta_n &= \max_r \{L(\eta_n(r+1)) - L(\eta_n(r))\} \\ &= \max_r \{2\eta_n(r+1)[1 - 2\eta_n(r+1)] - \eta_n(r)[1 - 2\eta_n(r)]\} \\ &\leq 2\frac{\epsilon_n}{d_n}(1 + 2\epsilon_n) = O(n^{-3/4} \log n) O(1) = O(n^{-3/4} \log n),\end{aligned}$$

as $\eta_n(r+1) - \eta_n(r) = \epsilon_n/d_n$.

We next verify that, with probability one, $K_n = O(n^{-3/4} \log n)$ as $n \rightarrow \infty$.

By the Borel-Cantelli Lemma it is sufficient to exhibit a constant L_1 for which

$$\sum_n P[K_n \geq L_1 n^{-3/4} \log n] < \infty.$$

First consider $P[|B_n(\eta_n(r))| \geq L_1 n^{-3/4} \log n]$. $|B_n(\eta_n(r))|$ can be written in the form $|U_n - EU_n|$ where

$$U_n = |L_n(\eta_n(r))| = \binom{n}{2}^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^n |I(X_i + X_j < 1, X_i \leq \eta_n(r) \text{ or } X_j \leq \eta_n(r))|,$$

a one-sample U statistic with associated, symmetric kernel

$$\psi(x_1, x_2) = I[x_1 + x_2 < 1, x_1 \leq \eta_n(r) \text{ or } x_2 \leq \eta_n(r)].$$

Take $z_n(r) = E\psi(X_1, X_2) = 2\eta_n(r)(1 - \eta_n(r)) \leq 2\epsilon_n$ and $\sigma_n^2(r) = \text{Var}\psi(X_1, X_2) = z_n(r)(1 - z_n(r)) \leq z_n(r) \leq 2\epsilon_n$. Note that $\max\{z_n(r), 1 - z_n(r)\} \leq 1$. Applying (6.2) we have

$$P[|U_n - EU_n| \geq L_1 n^{-3/4} \log n] \leq 2e^{-h}$$

where

$$h = \frac{[n/2] L_1^2 n^{-3/2} (\log n)^2}{2 \left(\sigma_n^2(r) + \frac{1}{3} \max \{z_n(r), 1 - z_n(r)\} \right)}.$$

Routine algebra then gives

$$h \geq \frac{L_1^2 n^{-1/2} (\log n)^2 e_n^{-1}}{8 (1 + e_n^{-1}) \left(\frac{n/2}{[n/2]} \right)}.$$

As $e_n \sim C_0 n^{-1/2} \log n$, there exist n_1, p_1 and p_2 such that for all $n \geq n_1$, $1 + e_n^{-1} \leq p_1$ and $(n/2)/[n/2] \leq p_2$. For $n \geq n_1$,

$$h \geq \left(\frac{L_1^2}{8p_1p_2} \right) \left(\frac{n^{-1/2} \log n}{e_n} \right) \log n.$$

Bound $(n^{-1/2} \log n)/e_n$ below by p_3 for $n \geq n_2$. Take L_1 such that $(L_1^2 p_3)/(8p_1p_2) = 2$. For $n \geq n_0 = \max\{n_1, n_2\}$ we have $h \geq 2 \log n$, implying that $e^{-h} \leq n^{-2}$. Then

$$P \left[K_n \geq L_1 n^{-3/4} \log n \right] \leq \sum_{r=1}^{d_n} P \left[|L_n(\eta_n(r))| \geq L_1 n^{-3/4} \log n \right] \leq d_n n^{-2}$$

for $n \geq n_0$. With $d_n \sim n^{-1/4}$,

$$\sum_{n=n_0}^{\infty} P \left[K_n \geq L_1 n^{-3/4} \log n \right] \leq \sum_{n=n_0}^{\infty} \frac{d_n}{n^2} < \infty.$$

As $H_n \leq K_n + \beta_n$, the proof is complete. ■

For $x \in [0, e_n]$,

$$L(x) = 2x(1 - x) = 2x - 2x^2 = 2x + O \left(n^{-3/2} (\log n)^2 \right),$$

therefore the following holds.

Corollary 6.1 *There exists L_2 such that*

$$P \left[\sup_{x \in I_n} \{|L_n(x) - 2x|\} \leq L_2 n^{-3/4} \log n \text{ a.a.} \right] = 1. \quad (6.3)$$

The next result shows that X^* , the contamination point, is sufficiently close to 0 to allow the use of Corollary 6.1 in passing from $L_n(X^*)$ to $2X^*$. As the sequence of critical values satisfies $(c_n - 1/2) = O(n^{-1/2})$, loosely applying Lemma 6.2 ($\zeta_1 = 1/12$), for some K ,

$$P \left[(c_n - V_n) \leq K n^{-1/2} \log n \text{ a.a.} \right] = 1. \quad (6.4)$$

Take Ω_0 to be the intersection of the sets implicit in (6.3) and (6.4), with $e_n \geq K n^{-1/2} \log n$; then $P[\Omega_0] = 1$. For n sufficiently large,

$$\begin{aligned} L_n(K_1 n^{-1/2} \log n) &\geq 2K n^{-1/2} \log n - L_2 n^{-3/4} \log n \\ &\geq K n^{-1/2} \log n \geq (c_n - V_n) I(V_n < c_n), \end{aligned}$$

i.e., with probability one, $X^* = O(n^{-1/2} \log n)$ as $n \rightarrow \infty$. The following can now be established.

Lemma 6.4 *With probability one,*

$$(c_n - V_n) I(V_n < c_n) - 2X^* = O(n^{-3/4} \log n), n \rightarrow \infty. \quad (6.5)$$

Proof: If $c_n \leq V_n$ the left-hand side of (6.5) is identically 0. Assume $c_n > V_n$. For any $\varepsilon > 0$

$$L_n(X^*) \geq c_n - V_n, \quad L_n(X^* - \varepsilon) < c_n - V_n. \quad (6.6)$$

For n sufficiently large, with probability one $X^* \in I_n = [0, e_n]$ with e_n chosen properly. Then for some $L_2 > 0$,

$$L_n(X^*) = 2X^* + (L_n(X^*) - 2X^*) \leq 2X^* + L_2 n^{-3/4} \log n, \quad (6.7)$$

$$L_n(X^* - \varepsilon) > 2X^* - 2\varepsilon - L_2 n^{-3/4} \log n. \quad (6.8)$$

Combining (6.7) and (6.8):

$$-2\varepsilon - L_2 n^{-3/4} \log n < (c_n - V_n) - 2X^* \leq L_2 n^{-3/4} \log n.$$

As ε is arbitrary, the result follows. ■

Results regarding the uniform convergence of the empirical distribution function in an $O(n^{-1/2} \log n)$ neighborhood of a specified quantile (here, the 0-th quantile) can be derived in the same fashion as the results stated in Chapter 5. In particular, with probability one,

$$\sup_{x \in I_n} \{|F_n(x) - x|\} = O(n^{-3/4} \log n), n \rightarrow \infty.$$

As, with probability one, $X^* \in I_n$ as $n \rightarrow \infty$, we have, with probability one,

$$\begin{aligned} F_n(X^*) &= X^* + O(n^{-3/4} \log n) \\ &= \frac{(c_n - V_n)}{2} + O(n^{-3/4} \log n), n \rightarrow \infty. \end{aligned}$$

The following is a direct consequence.

Theorem 6.1 *With probability one*

$$R_n = nF_n(X^*) = \frac{n(c_n - V_n)}{2} I(V_n < c_n) + O(n^{1/4} \log n), n \rightarrow \infty.$$

Hence $n^{-1/2}R_n \xrightarrow{\mathcal{L}} U$, U having the distribution of

$$12^{-1/2}(z^\alpha - Z)I(Z < z^\alpha),$$

where Z denotes the standard normal.

The Wilcoxon test is usually implemented in rank-sum form:

$$V_n^* = \frac{2}{n(n+1)} \sum_{i=1}^{n-1} \sum_{j=i+1}^n I\left(\frac{1}{2}(X_i + X_j) > 1/2\right).$$

Defining L_n^* analogously, we have

$$V_n^* = V_n + O(n^{-1}), \quad L_n^*(x) = L_n(x) + O(n^{-1}).$$

Therefore, the result of Theorem 6.1 holds for either version of the statistic.

A simulation was done to determine the accuracy of this approximation. For each of four sample sizes 5000 samples were drawn and forced to reject the $\alpha = 0.05$ test. The code was written in S-PLUS, a built-in function computed p -values using a normal approximation with a correction for continuity. Lower-most values were successively contaminated until $p < 0.05$. Results are in Table 6.1. A continuity correction of approximately +0.8 is suggested by the results.

6.4 Remarks

The development above can be adjusted to prove a corresponding result for the two-sided test.

The uniform integrability of the sequence $\{n^{-1/2}R_n\}$ is not established here (work in progress). The simulated values indicate that this result holds. Assuming ERR^* exists and is equal to the expected value of the limiting distribution, we have

Table 6.1: Simulated and Approximate Values for the Resistance to Rejection of the Wilcoxon Test Given as the Average Number of Contaminations Necessary to Force Rejection

n	<u>Resistance to Rejection</u>	
	Simulated	Approximate
30	3.449 (0.026)	2.634
60	4.575 (0.035)	3.725
120	6.092 (0.048)	5.267
240	8.273 (0.065)	7.450

Values in parentheses are estimated standard errors of the estimates.

$RR(S_n^+, V_n) = 1/\sqrt{3} \approx 0.577$. Oddly, this is almost exactly the ratio suggested by the studies cited in Chapter 5. In the case where the uncontaminated data most favor H_0 , the sign test (asymptotically) tolerates a contamination of $1/2$ of the entire sample; the Wilcoxon test tolerates a contamination of $1 - 1/\sqrt{2} \approx 0.293$ of the entire sample (a ratio of approximately 0.586).

Chapter 7

Summary

Quantile representation proved quite successful in leading to results regarding sign-scored methods of testing for trend. As a fortunate by-product, the structure of the representation of the vector of Mood statistics led to methods that improve the efficiency of pairwise testing methods when relative spacings are assumed and sample sizes are unequal . . . methods that extend to much broader class of score functions. The simulation study backs up the advisability of applying these methods. It also clearly supports the use of the $\bar{\chi}^2$ test in cases for which the entire space of isotonic (antitonic) response curves must be protected against.

Applying quantile representation methods to more general L_1 function fitting seems a reasonable extension.

The framework advanced in this work for evaluating test robustness illustrates that the statistical procedures commonly regarded as robust are not as resistant to contamination as other analyses indicate. The results derived for the sign-based methods clearly demonstrate a shortcoming. Slight, systematic measurement errors in one of the two samples can easily force a Mood test to falsely reject.

Work is still in progress establishing the uniform integrability result needed in Chapter 6. For the case of the Mann-Whitney two-sample rank statistic, W , the author has verified that the limiting distribution of R_n^x , the single sample contamination scheme, has the same form up to a constant as it does for the Mood statistic. The arguments are of the same nature of those supplied in Chapters 5 and 6. A

complete argument establishing the “closeness” of the small sample scheme to the optimal scheme has not been established. Provided this, as well as the uniform integrability condition, can be established, again we have $RR(W, M) = 1/\sqrt{3}$ (as in the one-sample case). The definitions of test breakdown in Rieder (1982) and Coakley and Hettmansperger (1992a), as well as developments of breakdown points for the corresponding estimates, all establish equal breakdown points for the Mann-Whitney and Mood procedures.

Developing expected resistance under the framework of a more general score function is of interest.

Bibliography

- Andrews, F. C. (1954). Asymptotic behavior of some rank tests for analysis of variance. *Annals of Mathematical Statistics*, *25*, 724–736.
- Arnold, S. F. (1981). *The Theory of Linear Models and Multivariate Analysis*. Wiley, New York, NY.
- Arnold, S. F. (1990). *Mathematical Statistics*. Prentice-Hall, Englewood Cliffs, NJ.
- Bahadur, R. R. (1966). A note on quantiles in large samples. *Annals of Mathematical Statistics*, *37*, 577–580.
- Bartholomew, D. J. (1959). A test of homogeneity for ordered alternatives. *Biometrika*, *46*, 36–48.
- Bartholomew, D. J. (1961). Ordered tests in the analysis of variance. *Biometrika*, *48*, 325–332.
- Billingsley, P. (1986). *Probability and Measure* (Second edition). Wiley, New York, NY.
- Chacko, V. J. (1963). Testing homogeneity against ordered alternatives. *Annals of Mathematical Statistics*, *34*, 945–956.
- Chase, G. R. (1974). On testing for ordered alternatives with increased sample size for a control. *Biometrika*, *61*, 569.
- Chernoff, H., & Savage, I. R. (1958). Asymptotic normality and efficiency of certain non-parametric test statistics. *The Annals of Mathematical Statistics*, *29*, 972–994.

- Coakley, C. W., & Hettmansperger, T. P. (1992a). Estimation Breakdown and Test Resistance in the Two-Sample Problem. In *Proceedings of the International Symposium on Nonparametric Statistics and Related Topics* Ottawa, ON, Canada.
- Coakley, C., & Hettmansperger, T. P. (1992b). Breakdown Bounds and Expected Test Resistance. *Journal of Nonparametric Statistics*, *1*, 267–276.
- Ghosh, J. K. (1971). A new proof of the Bahadur representation of quantiles and an Application. *Annals of Mathematical Statistics*, *42*, 1957–1961.
- He, X., Simpson, D. G., & Portnoy, S. L. (1990). Breakdown Robustness of Tests. *Journal of the American Statistical Association*, *85*, 446–452.
- Hettmansperger, T. P. (1984). *Statistical Inference Based on Ranks*. Wiley, New York.
- Hodges, Jr., J. L., & Lehmann, E. L. (1956). The efficiency of some nonparametric competitors of the t-test. *Annals of Mathematical Statistics*, *27*, 324–335.
- Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, *58*, 13–30.
- Jonckheere, A. R. (1954). A distribution-free k-sample test against ordered alternatives. *Biometrika*, *41*, 133–145.
- Mood, A. N. (1954). On the asymptotic efficiency of certain nonparametric two-sample tests. *The Annals of Mathematical Statistics*, *25*, 514–522.
- Noether, G. E. (1955). On a theorem of Pitman. *Annals of Mathematical Statistics*, *26*, 64–68.

- Pitman, E. J. G. (1949). Lecture Notes on Nonparametric Statistical Inference. Columbia University.
- Puri, M. L. (1964). Asymptotic efficiency of a class of c-sample tests. *The Annals of Statistics*, 35, 102–121.
- Puri, M. L. (1965). Some distribution-free k-sample rank tests of homogeneity against ordered alternatives. *Communications on Pure and Applied Mathematics*, 18, 51–63.
- Rao, C. R. (1973). *Linear Statistical Inference and Its Applications* (Second edition). Wiley, New York.
- Rieder, H. (1982). Qualitative robustness of rank tests. *Annals of Statistics*, 10, 205–211.
- Robertson, T., & Wright, F. T. (1982). Testing for ordered alternatives with increased precision in one of the samples. *Biometrika*, 69, 579–586.
- Robertson, T., & Wright, F. T. (1983). On approximation of the level probabilities and associated distributions in order restricted inference. *Biometrika*, 70, 597–606.
- Robertson, T., Wright, F. T., & Dykstra, R. L. (1988). *Order Restricted Statistical Inference*. Wiley, New York.
- Ryan, T. A. (1990). Lecture Notes on Statistical Computing. Pennsylvania State University.
- Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley, New York, NY.

- Shiraishi, T. (1982). Testing homogeneity against trend based on rank in a one-way layout. *Communications in Statistics*, 11.
- Shorack, G. R. (1967). Testing against ordered alternatives in model I analysis of variance; Normal theory and non-parametric. *Annals of Mathematical Statistics*, 38.
- Simpson, D. G. (1989). Hellinger deviance tests: efficiency, breakdown points, and examples. *Journal of the American Statistical Association*, 84, 107–113.
- Terpstra, T. J. (1953). The exact probability distribution of the T statistic for testing against trend and its normal approximation. *Proc. Sect. Sci. K. ned. Akad. Wet. (A)*, 56.
- Tryon, P. V., & Hettmansperger, T. P. (1973). A class of non-parametric tests for homogeneity against ordered alternatives. *The Annals of Statistics*, 1, 1061–1070.
- Uspensky, J. V. (1937). *Introduction to Mathematical Probability*. McGraw-Hill, New York.
- Ylvisaker, D. (1977). Test Resistance. *Journal of the American Statistical Association*, 72, 551–556.

VITA

SCOTT REESE PRESTON

Scott Reese Preston was born on June 6, 1958, in Evanston, Illinois. He graduated from East High School, Wausau, Wisconsin, in June of 1976. He earned a Bachelor of Science in Mathematics from the University of Wisconsin in August of 1982, and a Doctor of Philosophy in Statistics from The Pennsylvania State University.

Scott has been employed by the Kent School, Kent, Connecticut, Juniata College, Huntingdon, Pennsylvania, and is currently an Assistant Professor in the Department of Mathematics at the State University of New York, College at Oswego.