"One-Sided" Statistical Inference for a Multivariate Location Parameter

Inauguraldissertation
der Philosophisch-naturwissenschaftlichen Fakultät
der Universität Bern

vorgelegt von

Michael Vock
von Aarau und Wohlen AG

Leiter der Arbeit:
Prof. Dr. J. Hüsler
Institut für mathematische Statistik und Versicherungslehre

"One-Sided" Statistical Inference for a Multivariate Location Parameter

Inauguraldissertation
der Philosophisch-naturwissenschaftlichen Fakultät
der Universität Bern

vorgelegt von

Michael Vock
von Aarau und Wohlen AG

Leiter der Arbeit:
Prof. Dr. J. Hüsler
Institut für mathematische Statistik und Versicherungslehre

Von der Philosophisch-naturwissenschaftlichen Fakultät angenommen.

Bern, 25. Januar 2007

Der Dekan:
Prof. Dr. P. Messerli

Abstract

In this thesis, we approach the problem of "one-sided" statistical tests and corresponding confidence regions for a multivariate location parameter in the one-sample setting. By "one-sided", we mean that these methods incorporate information on the direction of a deviation from the null hypothesis that shall be detected. A strong emphasis is on nonparametric methods. Particular attention is also paid to the distinction between methods based on simple and those based on composite null hypotheses.

We propose several new test procedures and a graphical method for the comparison of different tests. This graphical method is useful for the assessment of the appropriateness of a test for a specific composite null hypothesis.

Finally, we use a highly flexible definition of confidence regions to derive results about their properties from those of the corresponding tests. Particularly, we obtain results about the shape of such regions.

Third printing, September 2007

Contents

1 Introduction and Outline **1**
1.1 Historical Background 1
1.2 Outline 2

2 One-Sided Multivariate Location Hypotheses **3**
2.1 Basic Multivariate One-Sided Alternatives 4
2.1.1 Direction Alternatives 4
2.1.2 Sector and Cone Alternatives 5
2.2 Simple vs. Composite Null Hypotheses 7
2.3 Hypotheses for Non-Inferiority Problems 8
2.4 Translations of the Hypotheses 10
2.5 Two- and Multi-Sample Problems 10

3 Properties of Tests **11**
3.1 Invariance Properties 11
3.1.1 Affine Transformations of p-Variate Data 11
3.1.2 Invariance of Tests 12
3.1.3 More General Transformation Classes 12
3.1.4 Transformations and One-Sided Tests 12
3.2 Unbiasedness 14
3.3 Cone Order Monotonicity 15
3.4 Assumptions on the Underlying Distribution 16

4 One-Sided Tests from the Literature **19**
4.1 One-Sample Problem 19
4.2 Two-Sample Problem 22
4.3 Multi-Sample Problem 23
4.4 Monotone Trend Problem 24
4.5 Combination of Univariate Tests 24
4.6 Related Problems 25

5 Union-Intersection and Intersection-Union Tests **26**
5.1 Componentwise Procedures for the $(2^p - 1)$-Orthant Alternative . 26
5.2 Componentwise Procedures for the One-Orthant Alternative . . . 27
5.3 Union-Intersection Tests 28
5.4 Intersection-Union Tests 28

6 Binomial Tests **29**
6.1 Half-Space 29
6.2 Cone/Opposite Cone 29
6.3 Estimated Cone Probability 30

7 Spatial Sign and Rank Tests **32**
7.1 Spatial Sign Test for an Unrestricted Alternative 32
7.2 Modification for a Direction Alternative 33
7.3 Adaptation to a Sector Alternative 35
7.4 Spatial Signed Rank Test 38

8 Tests Based on the Procedure by Randles (2000) **39**
8.1 Sign Test for an Unrestricted Alternative 39
8.2 Modification for a Direction Alternative 40
8.3 Adaptation to a Sector Alternative 42

9 "Hodges-Type" Tests **44**
9.1 Hodges' Bivariate Sign Test 44
9.2 Sign Test by Larocque and Labarre (2004) 44
9.3 Modification for a One-Orthant Alternative 45
9.4 Extension to a Convex Cone Alternative 46

10 Graphical Comparison Methods **51**
10.1 Acceptance and Rejection Regions 51
10.2 Power at a Fixed Distance from the Origin 54
10.3 Curve of Constant Power 58

11 Comparison of the Proposed Tests **63**
11.1 Uncorrelated Normal Case 64
11.1.1 Power at a Fixed Distance from the Origin 64
11.1.2 Curve of Constant Power 66
11.2 Effects of Correlation in the Normal Case 69
11.3 Heavy Tails: Bivariate $t(1)$ 69
11.4 "Exponential Radius" Distribution 71

12 Confidence Regions **73**
12.1 Connection with Hypothesis Tests 73
12.2 Shape of the Confidence Region for the Meta-Parameter $\boldsymbol{\gamma}$ 75
12.2.1 Translations of the Acceptance Region in $\mathbb{R}^p$ 75
12.2.2 Bounds for the Border of the Confidence Region 77
12.2.3 Bivariate Case 80
12.3 Confidence Regions for $\boldsymbol{\vartheta}$ Based on Simple Null Hypotheses . . . 82
12.4 Sharpened Confidence Regions for $\boldsymbol{\vartheta}$ 83

13 Example: Pulmonary Function Data **87**
13.1 Data 87
13.2 Tests 89
13.3 Confidence Regions 89

14 Discussion and Outlook **93**

Acknowledgements **95**

A Symmetry Centers and Expectations **96**
A.1 Interpretation of Symmetry Centers Using Expectations 96
A.2 Direction of the Expected Spatial Sign 97

B Underlying Results **101**
B.1 Asymptotic Distribution of Random Vectors 101
B.2 Inequalities for Binomial Probabilities 102
B.3 Minimum of Two Projections of an $\mathrm{N}_2(\mathbf{0}, \mathrm{I}_2)$ Random Vector . . . 103
B.4 Maximum of a Symmetric Binary Random Walk 104
B.5 Convex Cones 105

C Algorithm for the Estimation of the Curve of Constant Power **106**

Bibliography **108**

Index **115**

Chapter 1

Introduction and Outline

1.1 Historical Background

Elementary statistical textbooks provide many univariate data examples, and the statistical methods for their analysis are well-known, even to many non-statisticians. However, practically arising scientific questions tend to be more complicated – very often, more than one variable is of interest. It may be much more appropriate to analyze the information of several variables at once instead of doing separate analyses.

An early example of such a multivariate method can be found in the article by Hotelling (1931), in which the T^2 statistic was proposed as a multivariate generalization of Student's t for tests about a location parameter. Early nonparametric multivariate location tests were proposed by Hodges (1955) and Blumen (1958) (both for the bivariate case). The topic of such nonparametric tests has been receiving considerable attention during the last years; examples are the articles by Brown and Hettmansperger (1987), Randles (1989, 2000), Möttönen and Oja (1995), as well as Larocque, Tardif, and van Eeden (2000).

We will deal with one-sided statistical inference for a multivariate location parameter. By "one-sided", we mean that these methods incorporate information on the direction of a deviation from the null hypothesis that shall be detected. (An alternative hypothesis corresponding to such a test will also be called "restricted".) As we will see in Chapter 2, there are many more possibilities for the specification of such information in the multivariate case than in the univariate case.

Lehmann (1952) discussed general issues in testing one type of one-sided hypotheses. Kudô (1963) and Perlman (1969) made early proposals for multivariate one-sided tests under the assumption of normality. Among the first nonparametric procedures for such problems (in the two-sample case) are those by Bhattacharyya and Johnson (1970) and Chatterjee and De (1972). (Many further nonparametric one-sided location tests will be mentioned in Chapter 4.)

Most of the publications emphasize the concept of statistical tests, and not the dual concept of confidence regions. While confidence regions are immediately implied by the formulation of a (non-randomized) test, the practical aspects of such methods for multivariate one-sided problems do not seem to have received much attention.

1.2 Outline

We start with a chapter about different formulations of one-sided null and alternative hypotheses about a multivariate location parameter. Chapter 3 introduces several properties that may be desirable for the tests, e. g. invariance properties under transformation classes, but also validity under different assumptions on the distribution of the data. In Chapter 4, we give a review of multivariate tests with restricted alternatives from the literature.

The following chapters are focused on one-sample tests for cone (especially convex cone) alternatives. We begin with tests based on componentwise methods and so-called union-intersection and intersection-union tests as a related, but more general approach (Chapter 5). In Chapter 6, several nonparametric tests based on the number of observations in the alternative parameter region are discussed. Chapters 7 through 9 present specific nonparametric multivariate location tests proposed in the literature and modifications of these tests for one-sided problems.

We discuss two existing and one novel approach to graphical comparisons of different multivariate one-sided location tests in Chapter 10. We then apply our graphical approach to a selection of our own tests and of tests from the literature in Chapter 11.

In Chapter 12, we examine the connection between hypothesis tests and confidence regions. We use this connection to derive shape properties for different types of one-sided confidence regions for multivariate location parameters.

Finally, hypothesis tests and confidence regions are illustrated using an example from the literature in Chapter 13, and the main achievements of the thesis as well as possible extensions are discussed in Chapter 14.

Chapter 2

One-Sided Multivariate Location Hypotheses

The usual p-variate one-sample location problem can (up to translations, see Section 2.4) be formulated as

$$H_0 : \boldsymbol{\vartheta} = \mathbf{0} \quad \text{vs.}$$
$$H_1 : \boldsymbol{\vartheta} \neq \mathbf{0},$$

where $\boldsymbol{\vartheta}$ is a location parameter in $\mathbb{R}^p$. (More general alternatives than this "shift" alternative will not be considered here.)

The corresponding k-sample problem ($k \geq 2$) is

$$H_0 : \boldsymbol{\vartheta}_1 = \ldots = \boldsymbol{\vartheta}_k \quad \text{vs.}$$
$$H_1 : \boldsymbol{\vartheta}_i \neq \boldsymbol{\vartheta}_j \text{ for some } i, j \in \{1, \ldots, k\}.$$

In some cases, the possible range of parameters is only a subset of $\mathbb{R}^p$. As an example, some components may be restricted to be non-negative by theoretical considerations.

Furthermore, before looking at the data, one may already have an idea of the "direction" of the deviation of the location parameter from the null value, or it may even be the aim of a study to show a deviation into a specific "direction".

In all these cases, conventional tests with unrestricted alternatives are inappropriate. Their power may be unacceptably low, and they do not allow for showing a deviation into a specified direction. Thus special tests are needed that are able to deal with restricted alternatives. (The above-mentioned, quite different reasons for the usage of restricted alternatives will be returned to in Section 2.2.)

The usual restricted alternatives in the one-dimensional case are the well-known one-sided alternatives. It is not obvious how the concept of one-sided alternatives should be generalized to the multivariate case, so that careful consideration seems worthwhile. Emphasis will be on the one-sample location problem.

2.1 Basic Multivariate One-Sided Alternatives

In the one-dimensional one-sample setting, there are (up to translations) only two possible one-sided location (shift) alternatives:

$$H_1^+ : \ \vartheta > 0 \quad \text{and}$$
$$H_1^- : \ \vartheta < 0,$$

where the (simple) null hypothesis is

$$H_0 : \ \vartheta = 0.$$

The generalization to the multivariate case can be done in different ways. For the ease of formulation and visualization, we focus on the two-dimensional setting at first.

2.1.1 Direction Alternatives

We can interpret the univariate one-sided alternatives as the (only) two directions from the origin on the real line. In the bivariate case, if $\boldsymbol{\vartheta} = (\vartheta_1, \vartheta_2)^{\mathrm{T}}$ is the (now vector-valued) location parameter again, we can e. g. consider restricted alternatives of the following form:

$$H_1^{\varphi} : \ \langle \boldsymbol{\vartheta}, \boldsymbol{u} \rangle > 0,$$

where $\langle \cdot, \cdot \rangle$ denotes a scalar product (typically, the standard Euclidean scalar product) and $\boldsymbol{u} = (\cos\varphi, \sin\varphi)^{\mathrm{T}}$ is the unit vector at the angle φ; cf. Figure 2.1. Thus a projection of $\boldsymbol{\vartheta}$ is investigated. We call this type of alternative hypothesis a *direction alternative*.

The generalization of the direction alternative approach to the general p-variate case is straightforward.

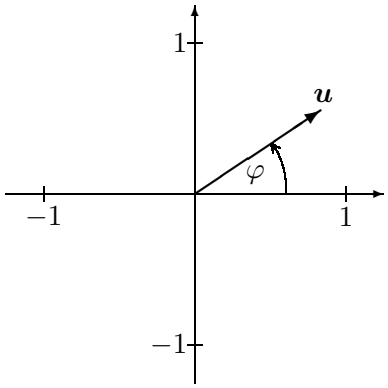

Figure 2.1: Specification of a direction alternative.

2.1.2 Sector and Cone Alternatives

Another possibility is to generalize the idea of partitioning the real line into the positive half-line $\mathbb{R}_+$ and the negative half-line $\mathbb{R}_-$. The simplest possibility in the bivariate case would be to consider half-planes, which is closely related to the approach using projections. It is perhaps more promising to use the more general concept of partitioning the plane into two sectors, i. e. to ask if the location parameter $\boldsymbol{\vartheta}$ belongs to a specified (infinite) sector between the angles φ_1 and φ_2 (assuming, without loss of generality, $\varphi_1 < \varphi_2$); cf. Figure 2.2 (a). The directions can equivalently be given by the unit vectors $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$ again. Formally, we can specify such a restricted alternative as

$$H_1^{\varphi_1,\varphi_2} : \boldsymbol{\vartheta} \neq \mathbf{0};\ \exists\, \psi \in [\varphi_1, \varphi_2] : \cos\psi = \frac{\vartheta_1}{\|\boldsymbol{\vartheta}\|},\ \sin\psi = \frac{\vartheta_2}{\|\boldsymbol{\vartheta}\|}.$$

(Sometimes, we will also use the open interval (φ_1, φ_2) instead of the closed interval.) We call this type of alternative a *sector alternative*.

This approach includes the case of half-planes (cf. Figure 2.2 (b), with a boundary slightly different from that in the direction alternative above). It also includes the quadrant alternatives, which are of considerable interest in many applications and have been discussed in several articles (e. g. Kudô, 1963; Chinchilli and Sen, 1981; Chatterjee and De, 1972; Boyett and Shuster, 1977; Park, Na, and Desu, 2001). A typical example for a one-quadrant alternative (cf. Figure 2.2 (c)) is the comparison of treatments where a new treatment should be superior to an old one in two criterions (possibly with equality in one criterion). In some cases, a treatment will be interesting for further research if it is superior in at least one of two criterions, regardless of a possible inferiority in the other criterion. This corresponds to the three-quadrant alternative (cf. Figure 2.2 (d)).

We could think of much more general forms of dividing up the plane. However, the sectors have the advantage of remaining unchanged if both coordinate axes are scaled by the same positive factor. The class of sector alternatives is even closed with respect to affine transformations. From a practical point of view, especially in a context where nonparametric tests are desirable, it is also often difficult to specify more than the directions that we expect for a change of the location parameter.

For the case of $p > 2$ dimensions, several authors (e. g. Perlman, 1969; Silvapulle, Silvapulle, and Basawa, 2002) have considered *positively homogeneous sets* or *cones* (in a very general sense of the word) as alternative regions, i. e. sets $C \subset \mathbb{R}^p$ with the property that $\boldsymbol{x} \in C$ implies $a\boldsymbol{x} \in C$ for all positive real numbers a. The most widely-used positively homogeneous sets are the *convex cones*. However, neither of these terms is an exact generalization of the concept of sectors to the multivariate case: In $\mathbb{R}^2$, a sector with an angle of more than π is not a convex cone, while e. g. any union of two sectors is a positively homogeneous set but not necessarily a sector. We can define a true generalization of a sector

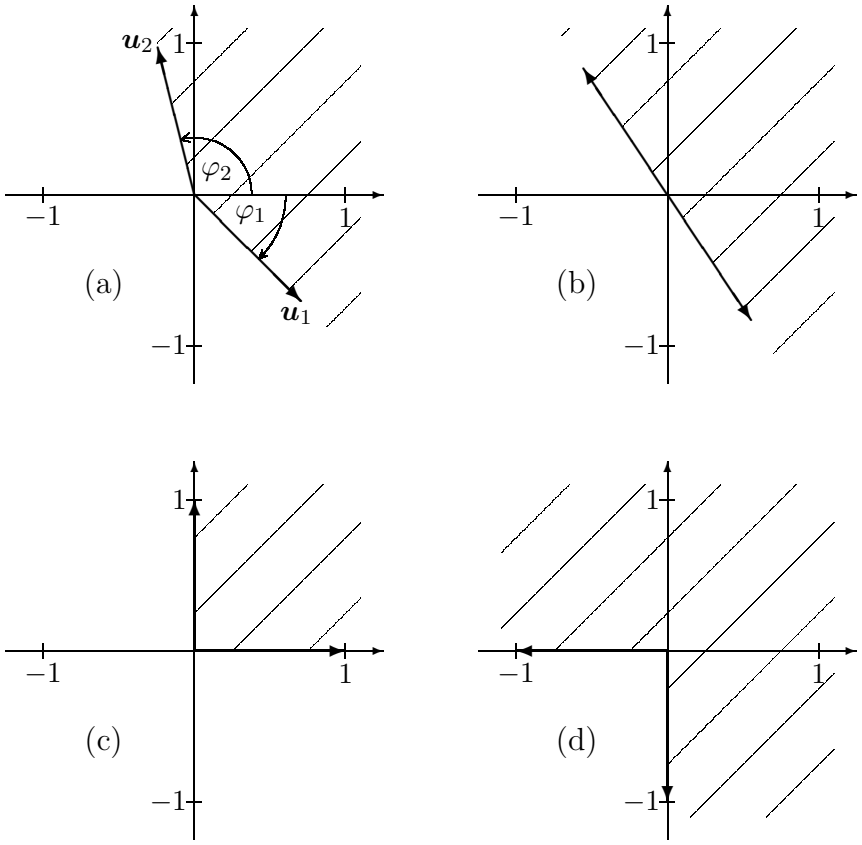

Figure 2.2: (a) Specification of a sector alternative; (b) Special sector alternative with $\varphi_2 = \varphi_1 + \pi$, i. e. half-plane; (c) Special sector alternative with $\varphi_1 = 0, \varphi_2 = \frac{\pi}{2}$, i. e. first quadrant; (d) Special sector alternative with $\varphi_1 = -\frac{\pi}{2}, \varphi_2 = \pi$, i. e. first, second, and fourth quadrant.

by requiring a positively homogeneous set $C \subset \mathbb{R}^p$ to have a simply connected intersection with the unit sphere S^{p-1}. We call such a set C an *s-simply-connected cone*. In many cases, it may be sensible to restrict attention to polyhedral cones, which can be specified by linear inequalities. Again, an important special case for hypotheses based on componentwise criteria are the *orthants*, which are the multivariate analogues of the quadrants.

In Section 2.3, we will deal with some modifications of the one-orthant alternative that do not fit into the framework of this section.

2.2 Simple vs. Composite Null Hypotheses

In the preceding section, we have written the null hypothesis as $H_0 : \boldsymbol{\vartheta} = \mathbf{0}$, i. e. as a simple (point) hypothesis. In the univariate case, for the tests widely used, it is e. g. irrelevant whether $H_0 : \vartheta = 0$ or $H_0 : \vartheta \leq 0$ is tested against $H_1 : \vartheta > 0$: If the test keeps its level for $H_0 : \vartheta = 0$, it also keeps its level for the composite null hypothesis. In the multivariate case, however, such a difference may be important, and, depending on the context, a composite null hypothesis may be more appropriate.

Let $H_1 : \boldsymbol{\vartheta} \in \Theta_1$ be the multivariate restricted alternative considered, e. g. with a positively homogeneous set for Θ_1. If we only consider a restricted alternative in order to enhance the power for $\boldsymbol{\vartheta} \in \Theta_1$ (because of some a priori conjecture) and if the aim is to show that $\boldsymbol{\vartheta}$ is different from $\mathbf{0}$, we can use the simple null hypothesis $\boldsymbol{\vartheta} = \mathbf{0}$. For $\boldsymbol{\vartheta} \in \mathbb{R}^p \setminus (\{\mathbf{0}\} \cup \Theta_1)$, we do not have to consider the level or the power in this case. The same trivially applies if the set $\Theta \subset \mathbb{R}^p$ of theoretically possible parameter values does not contain any point outside of $\{\mathbf{0}\} \cup \Theta_1$.

If, in contrast, $\{\mathbf{0}\} \cup \Theta_1$ is a true subset of Θ and if we want to show a deviation into a specific direction (the region Θ_1), we have to consider a rejection of the null hypothesis for $\boldsymbol{\vartheta} \in \Theta \setminus (\{\mathbf{0}\} \cup \Theta_1)$ as a type I error. In this case, tests with unrestricted alternatives do not only have poor power, they even do not necessarily respect the specified level. The appropriate hypotheses are then $H_0 : \boldsymbol{\vartheta} \in \Theta \setminus \Theta_1$ vs. $H_1 : \boldsymbol{\vartheta} \in \Theta_1$, i. e. the null hypothesis is composite. A typical example is that an amelioration (in some sense) in a treatment group with respect to a control group has to be shown. Berger (1982) mentions the example of a product that should be shown to meet all of several standards. In such cases, a composite null hypothesis is needed.

By "composite null hypothesis", we will usually mean the largest possible composite null hypothesis, i. e. $H_0 : \boldsymbol{\vartheta} \in \Theta_0 = \Theta \setminus \Theta_1$; intermediate cases will be indicated explicitly. Often, Θ will be taken to be $\mathbb{R}^p$ for simplicity.

The case where H_0 is composite is more appropriate for many applications, but also more difficult to handle than the case of a simple null hypothesis because the level has to be met for every $\boldsymbol{\vartheta} \in \Theta_0 = \Theta \setminus \Theta_1$. (This may be the reason why

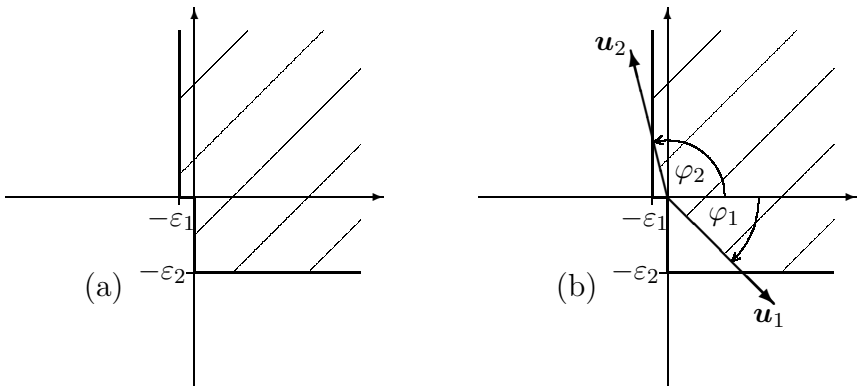

Figure 2.3: Special alternative hypotheses for non-inferiority problems: (a) Alternative region as in Bloch, Lai, and Tubert-Bitter (2001); (b) Intersection with a sector alternative.

it has only rarely been treated in the literature.) We will consider both types of null hypotheses occasionally. Namely, while we will use simple null hypotheses for the construction of some of our tests, we will place emphasis on the case of composite null hypotheses for assessing the performance of the tests.

One of the few articles that emphasize the importance of the distinction between the two types of null hypotheses is the one by Tang (1998).

2.3 Hypotheses for Non-Inferiority Problems

In univariate non-inferiority problems (also called "one-sided equivalence problems"), the intention is to show that some parameter ϑ is above or at least approximately equal to a specified value ϑ_0. Such a problem can be formalized as $H_0 : \vartheta \leq \vartheta_0 - \varepsilon$ against $H_1 : \vartheta > \vartheta_0 - \varepsilon$, where $\varepsilon > 0$ is some deviation that is considered irrelevant (e. g. Wellek, 2003). This is just a shifted version of the usual one-sided problem.

A more sophisticated formulation is based on an *indifference region* (Jennison and Turnbull, 1993) or *range of equivalence* (Freedman, Lowe, and Macaskill, 1983) between $\vartheta_0 - \varepsilon$ and $\vartheta_0 + \delta$ that separates the null and alternative parameter regions. If the true parameter is within this interval, it is not important whether we decide for H_0 or for H_1, and therefore we can neglect the rejection probabilities in this interval.

When we turn to the multivariate case, various approaches are possible: Bloch, Lai, and Tubert-Bitter (2001) propose to test $H_0 : \boldsymbol{\vartheta} \in \mathbb{R}^p \setminus \Theta_1$ against $H_1 : \boldsymbol{\vartheta} \in \Theta_1$ with $\Theta_1 = \{\boldsymbol{\vartheta} : \vartheta_j > -\varepsilon_j,\ \forall\, j \in 1, \ldots, p, \text{ and } \exists\, j \in 1, \ldots, p : \vartheta_j > 0\}$. This alternative region Θ_1 is shown in Figure 2.3 (a) for the bivariate situation. Jennison and Turnbull (1993) also propose this kind of alternative as one possibility;

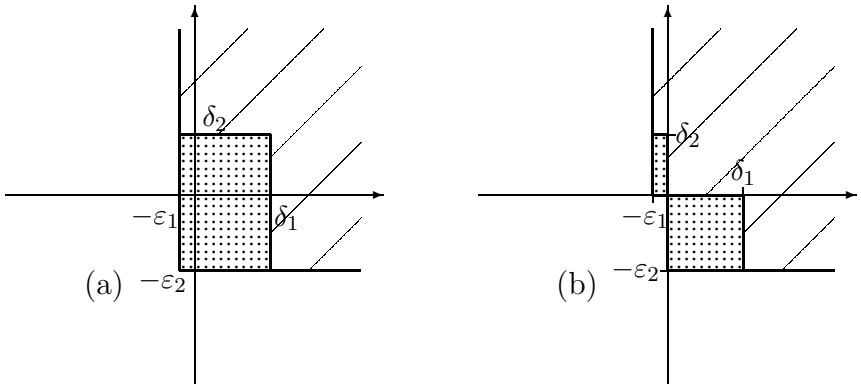

Figure 2.4: Non-inferiority alternative hypotheses (hatched) with indifference regions (dotted): (a) Cartesian product of componentwise indifference regions; (b) Indifference region reduced to the points where not all components have the same sign.

the other three possibilities proposed in their Figure 2 are shifted versions of an orthant alternative.

Conaway and Petroni (1996) formulate a sector alternative for the vector of the response rate and the toxicity rate in a tumor therapy. They allow for a trade-off between response and toxicity – e. g. a toxicity rate slightly above some reference value might be acceptable if the response rate is clearly above its reference value. This is a genuinely bivariate characterization of non-inferiority, in contrast to the ideas based on componentwise non-inferiority.

We can also combine the two approaches above: If we take the intersection of a sector alternative and the alternative from Figure 2.3 (a), the resulting alternative region is given in Figure 2.3 (b). It allows for a trade-off between the two parameters near the origin (in the sense proposed by Conaway and Petroni), but for each parameter, a strict lower bound is also given.

Based on the univariate indifference region approach, the most obvious multivariate version is the specification of the multivariate indifference region as a Cartesian product $(-\varepsilon_1, \delta_1] \times \ldots \times (-\varepsilon_p, \delta_p]$, as in Figure 2.4 (a). It may also be interesting to look at a slightly smaller indifference region that does not include parameter values having the same sign in all components, as shown in Figure 2.4 (b). As can easily be seen, the border between null and alternative region in both parts of Figure 2.3 goes through the indifference region in Figure 2.4 (b) if $\tan\varphi_1 = -\varepsilon_2/\delta_1$ and $\tan(\varphi_2 - \pi/2) = -\varepsilon_1/\delta_2$. This means that if some level α test for one of the problems without indifference region has some given power for every point in the alternative region, then this test is also of level α (or less) and has (at least) the same power for the situation in Figure 2.4 (b). If, in turn, a test is suitable when using the reduced indifference region from Figure 2.4 (b), it is also suitable for problems with the full indifference region in Figure 2.4 (a).

Tamhane and Logan (2004) propose a similar combination of hypotheses as in Figure 2.4 (a), but they incorporate the indifference region into the null hypothesis. This is just a shifted version of the situation in Figure 2.3 (a).

2.4 Translations of the Hypotheses

For notational simplicity, we have formulated all hypotheses in a version that is centered at $\mathbf{0}$. Of course, we also need tests for any hypothetical location parameter value other than the origin or for composite null and alternative regions that are constructed around some point (a meta-parameter $\boldsymbol{\gamma} \in \mathbb{R}^p$) other than the origin, i. e. for translated (shifted) versions of our various hypotheses.

We will maintain the simplified setting based on centered hypotheses in the following chapters. While doing so, however, we tacitly assume that tests for $\boldsymbol{\vartheta} \in \Theta_0(\boldsymbol{\gamma}) = \boldsymbol{\gamma} + \Theta_0(\mathbf{0})$ against $\boldsymbol{\vartheta} \in \Theta_1(\boldsymbol{\gamma}) = \boldsymbol{\gamma} + \Theta_1(\mathbf{0})$ are derived by application of a test for $\boldsymbol{\vartheta} \in \Theta_0(\mathbf{0})$ against $\boldsymbol{\vartheta} \in \Theta_1(\mathbf{0})$ to the translated data $\boldsymbol{X}_1 - \boldsymbol{\gamma}, \ldots, \boldsymbol{X}_n - \boldsymbol{\gamma}$, for all $\boldsymbol{\gamma} \in \Gamma$, where $\Gamma \subset \mathbb{R}^p$ is the set of possible meta-parameters. We also maintain the short notation $\Theta_0 = \Theta_0(\mathbf{0})$ and $\Theta_1 = \Theta_1(\mathbf{0})$.

For the construction of confidence regions in Chapter 12, we will have to return to the general setting and the notation with a meta-parameter $\boldsymbol{\gamma}$.

2.5 Two- and Multi-Sample Problems

For the two-sample problem, we can formulate hypotheses analogously to the one-sample case if we consider the difference of the location parameters, $\boldsymbol{\vartheta}_2 - \boldsymbol{\vartheta}_1$.

Similarly, in the multi-sample case, we can consider differences between location parameters of consecutive samples to formulate hypotheses. As in the univariate case, we may postulate a "strictly positive" difference (in the sense of the one-sample alternative) for at least one such comparison, and a "nonnegative" difference (i. e. a zero difference is also allowed) for the others.

There is an additional one-sided problem that has no analogue in the unrestricted context: the test for a *monotone trend* in a sequence of random vectors. This can be regarded as a degenerate multi-sample problem where each sample consists of one observation only.

Chapter 3

Properties of Tests

3.1 Invariance Properties

In the univariate setting, the common nonparametric tests are invariant with respect to quite general classes of transformations of the data. E. g., tests based on the ranking of the data are obviously invariant under strictly increasing transformations. For unrestricted alternatives, there has been a variety of proposals for multivariate nonparametric location tests having desirable invariance properties under certain transformations; see e. g. Oja (1999) for a review of affine invariant methods. We give an overview of some possible invariance properties and discuss their relevance in the context of one-sided tests.

3.1.1 Affine Transformations of p-Variate Data

Definition 3.1.1. Let $\boldsymbol{x}$ be a p-variate (observation) vector, A a $p \times p$ matrix, and $\boldsymbol{b}$ a (fixed) p-variate vector.

(a) $\boldsymbol{x} \mapsto \boldsymbol{x} + \boldsymbol{b}$ is a *location transformation* or *translation*.

(b) $\boldsymbol{x} \mapsto \mathrm{A}\boldsymbol{x}$ is a *scale transformation* if A is nonsingular and diagonal, i. e. a diagonal matrix with non-zero diagonal elements.

(c) $\boldsymbol{x} \mapsto \mathrm{A}\boldsymbol{x}$ is a *component permutation* if each row and each column of A contains exactly one 1 and all other entries are 0.

(d) $\boldsymbol{x} \mapsto \mathrm{A}\boldsymbol{x}$ is a *rotation* if A is orthogonal with determinant 1.

(e) $\boldsymbol{x} \mapsto \mathrm{A}\boldsymbol{x}$ is an *orthogonal transformation* if A is orthogonal.

(f) $\boldsymbol{x} \mapsto \mathrm{A}\boldsymbol{x} + \boldsymbol{b}$ is an *affine transformation* if A is nonsingular. △

Obviously, any location and/or scale transformation is affine. Any component permutation is an orthogonal transformation, any rotation is also an orthogonal

transformation, and any orthogonal transformation is affine. Thus the class of affine transformations contains all the transformations in the above definition.

If n p-variate observations are written as row vectors of an $n \times p$ matrix X and each observation is transformed according to $\boldsymbol{x} \mapsto \mathrm{A}\boldsymbol{x} + \boldsymbol{b}$, this can be written as $\mathrm{X} \mapsto \mathrm{X}\mathrm{A}^{\mathrm{T}} + \mathrm{B}$, where B is an $n \times p$ matrix containing $\boldsymbol{b}^{\mathrm{T}}$ in each row, and the result is again an $n \times p$ matrix with the transformed observations in its rows.

3.1.2 Invariance of Tests

Definition 3.1.2. Let $\mathcal{T}$ be a class of transformations $T : \mathbb{R}^p \to \mathbb{R}^p$. Let S_{Θ_0,Θ_1} be a test statistic for $H_0 : \boldsymbol{\vartheta} \in \Theta_0$ vs. $H_1 : \boldsymbol{\vartheta} \in \Theta_1$, based on n p-variate observations $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$, with $\Theta_0 \subset \mathbb{R}^p, \Theta_1 \subset \mathbb{R}^p \setminus \Theta_0$.

S is *$\mathcal{T}$-invariant* if, for all $T \in \mathcal{T}$ and $\boldsymbol{X}_i \in \mathbb{R}^p, i = 1, \ldots, n$,

$$S_{\Theta_0,\Theta_1}(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n) = S_{T(\Theta_0),T(\Theta_1)}(T(\boldsymbol{X}_1), \ldots, T(\boldsymbol{X}_n)). \qquad \triangle$$

Application of this definition to the transformation classes in Definition 3.1.1 leads to the concepts of translation invariance, scale invariance, component permutation invariance, rotation invariance, orthogonal invariance, and affine invariance, respectively.

3.1.3 More General Transformation Classes

In the univariate case, invariance of tests can not only be postulated for affine (i. e. linear) transformations. Therefore, in the generalization to the multivariate case, we can try to preserve invariance under more general classes of transformations than the affine ones.

Of course, we can consider *componentwise strictly monotone* transformations, which can be approximated by application of bicontinuous transformations to each component (*componentwise homeomorphisms*). As a class of "genuinely multivariate" transformations resembling the univariate concept of strict monotonicity, *bicontinuous transformations (homeomorphisms)* can also be considered.

Figure 3.1 gives an overview of the inclusions between some of the transformation classes discussed here and in Definition 3.1.1.

3.1.4 Transformations and One-Sided Tests

Under affine transformations, rays are mapped to rays, and positively homogeneous sets are mapped to positively homogeneous sets if the origin is moved appropriately. Therefore, it is adequate to ask whether a test for a positively homogeneous alternative is invariant under affine transformations (or some subclass of affine transformations). Similar considerations apply for componentwise strictly monotone transformations in the context of orthant alternatives. If we

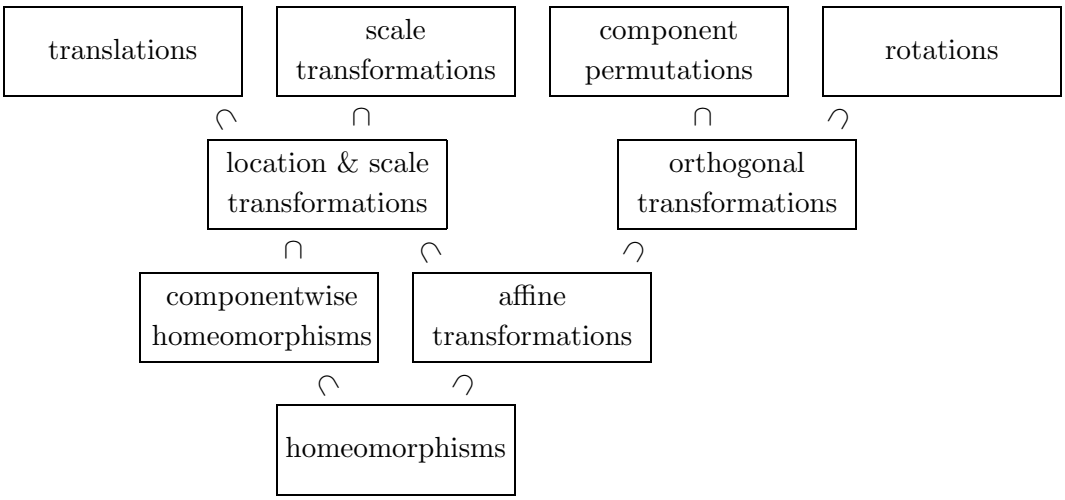

Figure 3.1: Inclusions between the different classes of transformations of multivariate data.

require affine invariance and invariance under componentwise strictly monotone transformations, we reach a high degree of independence from the way the data is represented in the variables.

However, these invariance requirements substantially constrain the choice of tests, and they are unnecessary for many multivariate applications. Namely, we can use less stringent requirements if we restrict our attention to special (but frequent) types of data and hypotheses:

For a direction alternative, a half-space is associated with the hypothetical direction, the border being the hyperplane that is orthogonal to the given direction. This orthogonality property is generally not preserved under affine (or scale) transformations, and therefore, invariance under such transformations may not be a sensible requirement. (Nevertheless, such a requirement may be sensible if the orthogonality, i. e. the scalar product, is defined in a coordinate system that depends on the distribution of the data.) Orthogonal transformations do not change orthogonality, and tests for direction alternatives may sensibly be investigated for the respective invariance properties. If we examine genuinely spatial data (i. e. data with all variables measured in the same units and with an arbitrary orientation of the coordinate system), orthogonal invariance is an obvious requirement, while e. g. scale changes are only sensible if they are applied to all variables simultaneously.

Alternatives that are defined using componentwise criteria (such as orthant alternatives and most of the alternatives from Section 2.3) do not usually stay within their class under more general transformations than componentwise homeomorphisms. Such alternatives are most interesting in the case of data that is not

genuinely spatial, where the variables represent an arbitrary collection of measurements typically measured in different units. But in this case, there is hardly any justification for applying rotations or even general affine transformations.

By construction (see Section 2.4), the tests proposed in the following chapters will be translation invariant. Most sensible tests will also be invariant under permutations of the components. Note that even a test for hypotheses with different criteria for the different components may well be invariant under permutations of the components since the transformation is also applied to the hypotheses for assessing invariance. A situation where this invariance property is not required could be the case of several univariate hypotheses of different importance.

The approach using bicontinuous transformations does not seem to be promising for one-sided problems: It is not difficult to find a bicontinuous transformation mapping two points that are in the same direction from the origin to points that are in opposite directions, while leaving the origin unchanged. (Imagine e. g. a bicontinuous transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ with $T((0,0)^{\mathrm{T}}) = (0,0)^{\mathrm{T}}$, $T((1,0)^{\mathrm{T}}) = (1,0)^{\mathrm{T}}$ and $T((2,0)^{\mathrm{T}}) = (-2,0)^{\mathrm{T}}$.) Thus, it is not appropriate to require a test to be invariant under such transformations if one-sided hypotheses formulated in terms of directions or positively homogeneous sets are considered.

3.2 Unbiasedness

In univariate problems, a test is usually required to be unbiased, i. e. its power function should be at most α for $\vartheta \in \Theta_0$ and at least α for $\vartheta \in \Theta_1$.

Lehmann (1952) investigates the situation where $H_0 : \vartheta_i \leq 0, i = 1, \ldots, p$ is tested against $H_1 : \exists\, i : \vartheta_i > 0$. He shows that the only test that has an analytic power function and that is unbiased for this problem is the trivial test with constant power α. We generalize this result to the general sector alternative (excluding only the half-space case) with the full composite null hypothesis. We do this in two steps:

Corollary 3.2.1. *Let a sample $(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)$ from a distribution from some family $(P_{\boldsymbol{\vartheta}})_{\boldsymbol{\vartheta} \in \mathbb{R}^2}$ be given. Let $\varphi : (\mathbb{R}^2)^n \to [0,1]$ be a test for $H_0 : \boldsymbol{\vartheta} \in \Theta_0$ vs. $H_1 : \boldsymbol{\vartheta} \in \Theta_1$, where $\Theta_1 = \mathbb{R}^2 \setminus \Theta_0$ is a sector with angle $\psi > \pi$. Assume that φ is an unbiased level α test and that it has an analytic power function.*

Then the test is trivial, i. e. the power of the test is constant: $\beta(\boldsymbol{\vartheta}) = \alpha$ under all $\boldsymbol{\vartheta} \in \mathbb{R}^2$.

Proof. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be an affine transformation such that $T(\Theta_0)$ is the negative quadrant. Define transformed observations $\tilde{\boldsymbol{X}}_i = T(\boldsymbol{X}_i)$, which have the distribution $\tilde{P}_{\tilde{\boldsymbol{\vartheta}}} = P_{T^{-1}(\tilde{\boldsymbol{\vartheta}})} \circ T^{-1}$.

$\tilde{\varphi}(\tilde{\boldsymbol{X}}_1, \ldots, \tilde{\boldsymbol{X}}_n) = \varphi(T^{-1}(\tilde{\boldsymbol{X}}_1), \ldots, T^{-1}(\tilde{\boldsymbol{X}}_n))$ can be used to test $\tilde{H}_0 : \tilde{\boldsymbol{\vartheta}} \in \tilde{\Theta}_0$ vs. $\tilde{H}_1 : \tilde{\boldsymbol{\vartheta}} \in \tilde{\Theta}_1$ with $\tilde{\Theta}_j = T(\Theta_j), j = 0, 1$. If $\tilde{\beta}$ denotes the power function of $\tilde{\varphi}$, then $\tilde{\beta}(T(\boldsymbol{\vartheta})) = \beta(\boldsymbol{\vartheta})$.

For a sample from a distribution from the family $(\tilde{P}_{\tilde{\boldsymbol{\vartheta}}})_{\tilde{\boldsymbol{\vartheta}}\in\mathbb{R}^2}$, $\tilde{\varphi}$ is therefore also an unbiased level α test for $\tilde{H}_0$ vs. $\tilde{H}_1$ and has an analytic power function. Because $\tilde{\Theta}_0 = T(\Theta_0)$ is the negative quadrant, Lehmann's result can be applied, and $\tilde{\beta}(\tilde{\boldsymbol{\vartheta}}) = \alpha, \forall\, \tilde{\boldsymbol{\vartheta}} \in \mathbb{R}^2$. But this implies also $\beta(\boldsymbol{\vartheta}) = \alpha, \forall\, \boldsymbol{\vartheta} \in \mathbb{R}^2$. □

Corollary 3.2.2. *Let a sample $(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)$ from a distribution from some family $(P_{\boldsymbol{\vartheta}})_{\boldsymbol{\vartheta}\in\mathbb{R}^2}$ be given. Let $\varphi : (\mathbb{R}^2)^n \to [0,1]$ be a test for $H_0 : \boldsymbol{\vartheta} \in \Theta_0$ vs. $H_1 : \boldsymbol{\vartheta} \in \Theta_1$, where $\Theta_1 = \mathbb{R}^2 \setminus \Theta_0$ is a sector with angle $\psi < \pi$. Assume that φ is an unbiased level α test and that it has an analytic power function.*

Then the test is trivial, i. e. the power of the test is constant: $\beta(\boldsymbol{\vartheta}) = \alpha$ under all $\boldsymbol{\vartheta} \in \mathbb{R}^2$.

Proof. Define $\tilde{\varphi}(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n) = 1 - \varphi(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)$, and let its power function be $\tilde{\beta}$. Trivially, $\tilde{\beta}(\boldsymbol{\vartheta}) = 1 - \beta(\boldsymbol{\vartheta})$. Therefore, $\tilde{\varphi}$ is an unbiased level $1-\alpha$ test for $\tilde{H}_0 : \boldsymbol{\vartheta} \in \Theta_1$ vs. $\tilde{H}_1 : \boldsymbol{\vartheta} \in \Theta_0$, and its power function is analytic. Due to Corollary 3.2.1, $\tilde{\beta}(\boldsymbol{\vartheta}) = 1 - \alpha, \forall\, \boldsymbol{\vartheta} \in \mathbb{R}^2$, and it follows that $\beta(\boldsymbol{\vartheta}) = \alpha, \forall\, \boldsymbol{\vartheta} \in \mathbb{R}^2$. □

From these results, we can see that strict unbiasedness will not be a useful criterion in most bivariate (and presumably also higher dimensional) problems, as long as the null hypothesis is composite. However, it is still desirable to have a "nearly unbiased" test. One method for assessing the bias of a test will be given in Section 10.3.

3.3 Cone Order Monotonicity

Cohen and Sackrowitz (1998) propose that tests for cone alternatives should satisfy the cone order monotonicity property with respect to Θ_1 and/or its *positive dual* $\Theta_1^* = \{\boldsymbol{a} : \boldsymbol{a}^{\mathrm{T}}\boldsymbol{\vartheta} \geq 0\ \forall\, \boldsymbol{\vartheta} \in \Theta_1\}$. Their definition of cone order monotonicity only applies to a test based on a single p-variate statistic. When we look at a general test based on a sample of n p-variate observations, at least two generalizations are possible; we introduce these in parts (b) and (c) of the following definition.

Definition 3.3.1. Let $C \subset \mathbb{R}^p$ be a convex cone.

(a) A function $f : \mathbb{R}^p \to \mathbb{R}$ is *cone order monotone with respect to C* if

$$f(\boldsymbol{x}) \leq f(\boldsymbol{x} + \boldsymbol{\delta}) \quad \forall\, \boldsymbol{x} \in \mathbb{R}^p, \boldsymbol{\delta} \in C.$$

(b) A function $f : (\mathbb{R}^p)^n \to \mathbb{R}$ is *cone order monotone in the sample with respect to C* if

$$f(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) \leq f(\boldsymbol{x}_1 + \boldsymbol{\delta}, \ldots, \boldsymbol{x}_n + \boldsymbol{\delta}) \quad \forall\, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \in \mathbb{R}^p, \boldsymbol{\delta} \in C.$$

(c) A function $f : (\mathbb{R}^p)^n \to \mathbb{R}$ is *cone order monotone in each observation with respect to* C if

$$f(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) \leq f(\boldsymbol{x}_1 + \boldsymbol{\delta}_1, \ldots, \boldsymbol{x}_n + \boldsymbol{\delta}_n) \quad \forall\, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \in \mathbb{R}^p, \\ \boldsymbol{\delta}_1, \ldots, \boldsymbol{\delta}_n \in C. \qquad \triangle$$

It is obvious that cone order monotonicity in each observation implies cone order monotonicity in the sample.

For a discussion of the adequacy of requiring cone order monotonicity, see e. g. Perlman and Chaudhuri (2004) and Cohen and Sackrowitz (2004).

Cohen and Sackrowitz (1998) give a method for constructing cone order monotone tests from tests for unrestricted alternatives by enlarging the acceptance region (and then determining the level of the new test). The authors always use the smallest superset of the acceptance region such that the monotonicity condition is fulfilled. However, one could also use a larger superset of the same acceptance region, which would lead to a test with even smaller level. The resulting family of tests might be more appropriate for certain composite null hypotheses.

3.4 Assumptions on the Underlying Distribution

For a test to be valid, the distribution of the observations has to belong to a certain parametric family in the case of parametric tests. For nonparametric one-sample tests, we typically have to assume some kind of symmetry of the distribution.

Multivariate distributions can belong to several symmetry classes:

Definition 3.4.1. Let $\boldsymbol{X}$ be a p-variate random vector. The distribution $F_{\boldsymbol{X}}$ of $\boldsymbol{X}$ is

(a) *spherically symmetric (with respect to* $\mathbf{0}$*)* if, for every orthogonal $p \times p$ matrix A, $\boldsymbol{X} \stackrel{\mathrm{d}}{=} \mathrm{A}\boldsymbol{X}$;

(b) in the *spherical directions class (with respect to* $\mathbf{0}$*)* if there exists a random vector $\boldsymbol{Y}$ with a spherically symmetric distribution (with respect to $\mathbf{0}$) such that $\frac{\boldsymbol{X}}{\|\boldsymbol{X}\|} \stackrel{\mathrm{d}}{=} \frac{\boldsymbol{Y}}{\|\boldsymbol{Y}\|}$;

(c) *elliptically symmetric (with respect to* $\mathbf{0}$*)* if there exists a nonsingular $p \times p$ matrix B such that the distribution of $\mathrm{B}\boldsymbol{X}$ is spherically symmetric;

(d) in the *elliptical directions class (with respect to* $\mathbf{0}$*)* if there exists a random vector $\boldsymbol{Y}$ with an elliptically symmetric distribution (with respect to $\mathbf{0}$) such that $\frac{\boldsymbol{X}}{\|\boldsymbol{X}\|} \stackrel{\mathrm{d}}{=} \frac{\boldsymbol{Y}}{\|\boldsymbol{Y}\|}$;

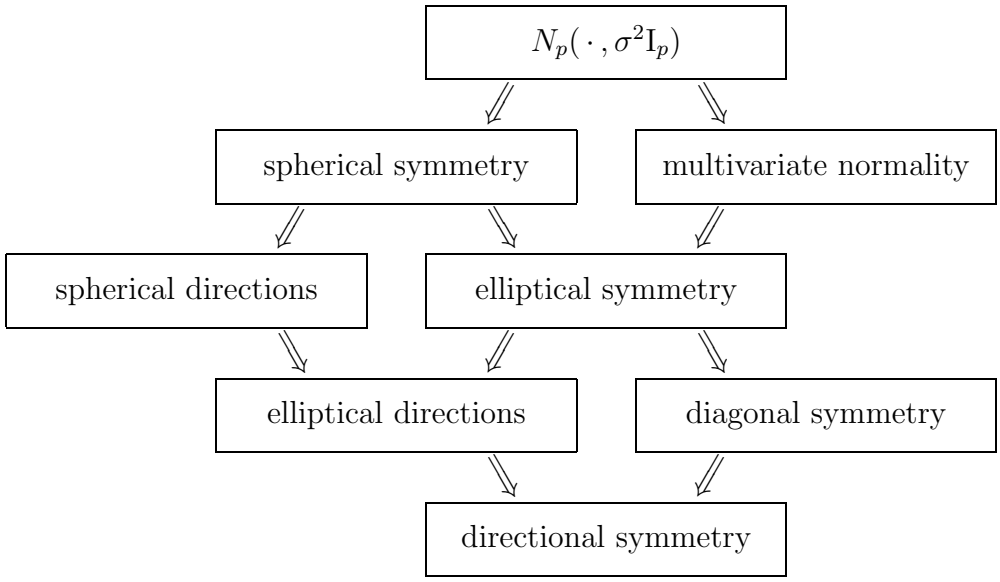

Figure 3.2: Implications between the different types of symmetry of multivariate distributions.

(e) *diagonally symmetric (with respect to* $\mathbf{0}$*)* if $\boldsymbol{X} \stackrel{\mathrm{d}}{=} -\boldsymbol{X}$;

(f) *directionally symmetric (with respect to* $\mathbf{0}$*)* if $\frac{\boldsymbol{X}}{\|\boldsymbol{X}\|} \stackrel{\mathrm{d}}{=} -\frac{\boldsymbol{X}}{\|\boldsymbol{X}\|}$.

For $\boldsymbol{\vartheta} \in \mathbb{R}^p$, $F_{\boldsymbol{X}}$ is *symmetric* (in any of the above senses) *with respect to* $\boldsymbol{\vartheta}$ if $\boldsymbol{X} - \boldsymbol{\vartheta}$ is symmetric (in the same sense) with respect to $\mathbf{0}$. △

Spherical symmetry trivially implies the spherical directions property (with $\boldsymbol{Y} = \boldsymbol{X}$) and elliptical symmetry (with the identity matrix I_p for B). As a consequence of the latter implication, a distribution in the spherical directions class is also in the elliptical directions class. Elliptical symmetry implies the elliptical directions property (with $\boldsymbol{Y} = \boldsymbol{X}$) and diagonal symmetry (with $\mathrm{A} = -\mathrm{I}_p \Rightarrow \mathrm{B}\boldsymbol{X} \stackrel{\mathrm{d}}{=} -\mathrm{B}\boldsymbol{X} \stackrel{\mathrm{d}}{=} \mathrm{B}(-\boldsymbol{X}) \Rightarrow \boldsymbol{X} \stackrel{\mathrm{d}}{=} -\boldsymbol{X}$). Diagonal symmetry and the elliptical directions property each again trivially imply directional symmetry. These relationships and the obvious relationships with the special case of multivariate normal distributions are illustrated in Figure 3.2, which is an extension of Figure 1 in Randles (2000).

If $\boldsymbol{X}$ has a probability density function $f_{\boldsymbol{X}}$, we can also characterize some of the above symmetry properties as follows:

- Spherical symmetry: $f_{\boldsymbol{X}}(\boldsymbol{x}) = g(\|\boldsymbol{x}\|)$ for some $g : [0, \infty) \to [0, \infty)$.
- Elliptical symmetry: $f_{\boldsymbol{X}}(\boldsymbol{x}) = g(\sqrt{x^{\mathrm{T}}\mathrm{B}^{\mathrm{T}}\mathrm{B}x})\,|\det \mathrm{B}|$, or, in a more habitual form, $f_{\boldsymbol{X}}(\boldsymbol{x}) = \tilde{g}(x^{\mathrm{T}}\Sigma^{-1}x)\,|\det \Sigma|^{-\frac{1}{2}}$, with $\Sigma = (\mathrm{B}^{\mathrm{T}}\mathrm{B})^{-1}$.

- Diagonal symmetry: $f_{\boldsymbol{X}}(\boldsymbol{x}) = f_{\boldsymbol{X}}(-\boldsymbol{x})$ for all $\boldsymbol{x} \in \mathbb{R}^p$.

Diagonal symmetry is often just called symmetry. Further, diagonal symmetry is also known as "central symmetry" or "reflected symmetry", and directional symmetry is also called "angular symmetry"; see e. g. Small (1990), Neuhaus and Zhu (1999).

The terms used here are mainly based on those used in Hettmansperger and McKean (1998) and Randles (2000). Randles's definition of the elliptical directions class for observations $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ is slightly less restrictive: He does not require the observations to be i. i. d., while we maintain the classical framework of a test based on an i. i. d. sample. The following definition by Neuhaus and Zhu (1999) is equivalent to the one given here: The distribution of $\boldsymbol{X}$ is in the elliptical directions class (with respect to $\boldsymbol{0}$) if there exists a matrix B such that $\frac{\mathrm{B}\boldsymbol{X}}{\|\mathrm{B}\boldsymbol{X}\|}$ is uniformly distributed on the unit sphere.

We have introduced the spherical directions class here in analogy to the elliptical directions class, and it is easily seen that an equivalent (and simpler) definition would be that $\frac{\boldsymbol{X}}{\|\boldsymbol{X}\|}$ is uniformly distributed on the unit sphere.

Chapter 4

One-Sided Tests from the Literature

There has been a variety of proposals for tests with "one-sided", "ordered", or "restricted" alternatives in multivariate settings. The emphasis in this survey is on nonparametric methods.

As the hypotheses for multivariate location tests with restricted alternatives are often formulated in terms of componentwise comparisons, the following notation will be useful, given two vectors $\boldsymbol{a} = (a_1, \ldots, a_p)^{\mathrm{T}}$ and $\boldsymbol{b} = (b_1, \ldots, b_p)^{\mathrm{T}}$:

$$\begin{aligned}
&\boldsymbol{a} \leq \boldsymbol{b} \Leftrightarrow a_i \leq b_i,\ i = 1, \ldots, p \\
&\boldsymbol{a} \overset{\exists}{\leq} \boldsymbol{b} \Leftrightarrow a_i \leq b_i,\ i = 1, \ldots, p, \text{ and } \exists\, i : a_i < b_i \\
&\boldsymbol{a} < \boldsymbol{b} \Leftrightarrow a_i < b_i,\ i = 1, \ldots, p
\end{aligned}$$

$\boldsymbol{a} \geq \boldsymbol{b}$, $\boldsymbol{a} \overset{\exists}{\geq} \boldsymbol{b}$, and $\boldsymbol{a} > \boldsymbol{b}$ will be used analogously.

4.1 One-Sample Problem

- Brown (1983) describes an "angle test", a rotation invariant bivariate analogue of a sign test based on the statistic

$$\sum_i \cos(\psi_i - \varphi),$$

  where ψ_i is the angle between the positive part of the first coordinate axis and the i-th observation, and φ is the hypothetical direction as in Figure 2.1. Brown uses a normal approximation to the distribution of the test statistic under the null hypothesis.

- In the introduction to their article, Larocque and Labarre (2004) announce to propose a test for the one-orthant alternative $H_1 : \boldsymbol{\vartheta} \overset{\exists}{\geq} \mathbf{0}$. However, the

conditionally distribution-free sign test proposed actually seems to be more appropriate for the alternative that at least one component of $\boldsymbol{\vartheta}$ is positive, without any restriction being imposed on the other components (bivariate case: three-quadrant alternative, Figure 2.2 (d)). The test is based on the supremum of univariate sign test statistics on projections of the data and will be presented in more detail in Section 9.2.

- Chinchilli and Sen (1981) give a test of $H_0 : \boldsymbol{\vartheta} = (\vartheta_1, \ldots, \vartheta_p)^{\mathrm{T}} = \mathbf{0}$ against $H_1 : \boldsymbol{\vartheta} \neq \mathbf{0}$, $\vartheta_i \geq 0$ for $i = 1, \ldots, a$, with $a \leq p$ fixed, where $\boldsymbol{\vartheta}$ is the parameter vector of a general linear model. They use the union-intersection principle (see Section 5.3). In the special case $a = p = 2$, the alternative hypothesis corresponds to Figure 2.2 (c).

- Silvapulle, Silvapulle, and Basawa (2002) develop a class of adaptive tests for the null hypothesis $\boldsymbol{\vartheta} = \mathbf{0}$ against the alternative $\boldsymbol{\vartheta} \in C \setminus \{\mathbf{0}\}$, where $\boldsymbol{\vartheta} \in \mathbb{R}^p$ is a parameter vector and C is a closed convex subset of $\mathbb{R}^p$ containing the origin (or, as a special case, a closed convex positively homogeneous set). This setting includes the half-plane and one-quadrant alternatives discussed for the bivariate case (Figure 2.2 (b), (c)). The approach of these authors is based on asymptotic considerations.

- Minhajuddin, Frawley, Schucany, and Woodward (2006) propose two bootstrap tests. The first one is for the simple null hypothesis $\boldsymbol{\vartheta} = \mathbf{0}$ against the alternative that $\boldsymbol{\vartheta}$ is in the positive orthant. It is based on the likelihood ratio test statistic used by Perlman (1969) in the multivariate normal context – see below.

  The second test proposed is for the same alternative, but for the composite null hypothesis consisting of the complement of the positive orthant. Resampling is done after centering the data in such a way that the mean is on the border of the null parameter region whenever the original mean was in the positive orthant.

- Parametric procedures: Kudô (1963) considers a p-variate normal population with mean $\boldsymbol{\vartheta}$ and known covariance matrix. He proposes a likelihood ratio test for $H_0 : \boldsymbol{\vartheta} = \mathbf{0}$ against $H_1 : \boldsymbol{\vartheta} \overset{\exists}{\geq} \mathbf{0}$. This alternative hypothesis corresponds to the bivariate situation in Figure 2.2 (c). Perlman (1969) extends this work to positively homogeneous sets as alternative regions and to the case of an unknown covariance matrix, giving upper and lower bounds on the null distribution of the test statistic. A discussion of these tests and related results is given in Section 4.6 of Robertson, Wright, and Dykstra (1988).

  An approximation to these likelihood ratio tests is given by Tang, Gnecco, and Geller (1989).

For a summary of early parametric approaches in the context of (convex) positively homogeneous sets, see Shapiro (1988).

Akkerboom (1990) proposes to test a simple null hypothesis against a polyhedral cone alternative by using a circular cone that approximates the original polyhedral cone.

A composite null hypothesis problem is investigated by Sasabuchi (1988). He considers a p-variate normal random vector with mean $\boldsymbol{\vartheta}$ and unknown covariance matrix. For given p-variate vectors $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_k$ satisfying certain conditions, he derives the likelihood ratio test of $H_0 : \boldsymbol{b}_i^{\mathrm{T}}\boldsymbol{\vartheta} \geq 0$ for all i with at least one equality against $H_1 : \boldsymbol{b}_i^{\mathrm{T}}\boldsymbol{\vartheta} > 0$ for all i. In this context, the above formulation of the null hypothesis is equivalent to $H_0 : \exists\, i : \boldsymbol{b}_i^{\mathrm{T}}\boldsymbol{\vartheta} \leq 0$. See also Berger (1989), where two more powerful modifications of the test are proposed, of which at least one has counterintuitive properties. Further similar modifications are proposed by McDermott and Wang (2002), assuming a known covariance matrix.

In order to test if the components have positive means (assuming a known covariance matrix Σ), Follmann (1996) proposes to use the likelihood ratio test statistic $X^2 = n\bar{\boldsymbol{X}}^{\mathrm{T}}\Sigma^{-1}\bar{\boldsymbol{X}}$, but to reject the null hypothesis if both X^2 exceeds the critical value for the level 2α and the sum of the componentwise means is positive. In the case of an unknown Σ, Hotelling's T^2 can be used instead of X^2. In the bivariate case, the alternative is similar to the one in Figure 2.2 (b) (with $\varphi_1 = -\pi/4$) or Figure 2.1 (with $\varphi = \pi/4$).

Cohen and Sackrowitz (1998) give a general procedure to derive tests for $H_0 : \boldsymbol{\vartheta} = \mathbf{0}$ vs. $H_1 : \boldsymbol{\vartheta} \in C \setminus \{\mathbf{0}\}$ (where C is a closed convex positively homogeneous set) from tests for unrestricted alternatives when $\boldsymbol{\vartheta}$ is the natural parameter of an exponential family. Their procedure enlarges the acceptance regions of unrestricted tests in order to ensure cone order monotonicity of the tests – see Section 3.3. The significance levels for these enlarged acceptance regions are not directly related to those of the unrestricted tests and therefore have to be determined from scratch.

Mudholkar, Kost, and Subbaiah (2001) propose a robustified test based on trimmed means.

Glimm, Srivastava, and Läuter (2002) test $H_0 : \boldsymbol{\vartheta} = \mathbf{0}$ against $H_1 : \boldsymbol{\vartheta} \gneq \mathbf{0}$. They give the exact null distribution of several versions of a test statistic under a normal distribution. In order to simplify the calculation of these test statistics, the convex polyhedral cone resulting from an affine transformation of the positive orthant is approximated by an orthant. According to the authors, it can be shown that the null distribution of the test statistics is valid for data from any elliptically symmetric distributions.

For the problem of showing that at least one component of the parameter is

positive, Perlman and Wu (2006) propose a test that is more powerful than the likelihood ratio and union-intersection tests in cases where some of the components are negative. They also provide a related test for showing that at least one component is positive while the other components are above some slightly negative value, as in Figure 2.3 (a) for the bivariate case.

A survey of other, mostly parametric approaches can be found in Sen and Silvapulle (2002).

4.2 Two-Sample Problem

- Bhattacharyya and Johnson (1970) propose a "layer rank test" for the bivariate case. They consider two independent random samples $(\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_{n_1})$ and $(\boldsymbol{Z}_{n_1+1}, \ldots, \boldsymbol{Z}_{n_1+n_2})$, where $\boldsymbol{Z}_i = (X_i, Y_i)^{\mathrm{T}}$ follows a continuous distribution F for $i = 1, \ldots, n_1$ and G for $i = n_1 + 1, \ldots, n_1 + n_2$, respectively. The test problem is

$$\begin{aligned} &H_0 : \ F \equiv G \quad \text{vs.} \\ &H_1 : \ F \not\equiv G, F(x,y) \geq G(x,y), \bar{F}(x,y) \leq \bar{G}(x,y) \text{ for all } (x,y), \end{aligned}$$

where $\bar{F}(x,y) := \mathrm{P}(X_1 \geq x, Y_1 \geq y)$. (Under the conditions of H_1, a random vector with cdf F is called *strongly stochastically smaller* than a random vector with cdf G.) Thus Bhattacharyya and Johnson discuss a more general location problem than the shift problem presented in Section 2.1.

The test statistic used is

$$\frac{1}{(n_1+n_2)^2}\left[\frac{n_1}{n_1+n_2}\sum_{i=n_1+1}^{n_1+n_2}\sum_{j=1}^{n_1+n_2} 1(X_i \geq X_j, Y_i \geq Y_j) - \frac{n_2}{n_1+n_2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_1+n_2} 1(X_i \geq X_j, Y_i \geq Y_j)\right],$$

on which a permutation test is performed. It is shown that the statistic is invariant under the group of bicontinuous transformations $g : \mathbb{R}^2 \to \mathbb{R}^2$ satisfying

$$\boldsymbol{z}_1 \geq \boldsymbol{z}_2 \Rightarrow g(\boldsymbol{z}_1) \geq g(\boldsymbol{z}_2),$$

where the inequalities are to be interpreted componentwise, as defined in the introduction to this chapter.

Johnson and Mehrotra (1972) compare this test statistic to a more elaborate one, also based on coordinatewise rankings.

Wei and Knuiman (1987) adapt the test to the situation with censored data.

- Chatterjee and De (1972) present a test of $H_0 : \boldsymbol{\vartheta} = \mathbf{0}$ vs. $H_1 : \boldsymbol{\vartheta} \overset{\sqsupseteq}{\geq} \mathbf{0}$ in the bivariate case, where $\boldsymbol{\vartheta}$ is the location difference between the two samples (two-sample analogue of Figure 2.2 (c)). Their test statistic is the supremum of a family of statistics based on the rankings within each component.

  Chatterjee (1984) gives some additional considerations of this approach.

- Boyett and Shuster (1977) consider the problem of testing whether a positive location difference is present in at least one component (as in Figure 2.2 (d)). They apply a permutation test to the maximum of all componentwise t statistics (for the case of two independent or dependent samples). A test based on the minimum of componentwise t statistics is given for the problem where a positive location difference in all components has to be shown (as in Figure 2.2 (c)).

- O'Brien (1984) proposes a test for the comparison of two or more multivariate samples, where the null hypothesis is that of no difference and the alternative states that one sample has higher values in all components of the location parameter. It is not specified which sample should have higher values. Thus in the bivariate case, the area for the alternative consists of the first and the third quadrant. This is a problem that does not directly fit into the concept of sector (or s-simply-connected cone) alternatives – one would have to consider unions of sectors, which are, of course, still positively homogeneous. O'Brien calculates the sum of the coordinatewise rankings for each observation. A usual univariate ANOVA can be applied to these sums. O'Brien compares this proposal to parametric approaches.

  A common framework for O'Brien's nonparametric test and several parametric tests is given by Bregenzer and Lehmacher (1998).

- Park, Na, and Desu (2001) construct a test for $H_0 : \boldsymbol{\vartheta} \leq \mathbf{0}$ vs. $H_1 : \exists\, i : \vartheta_i > 0$ (Figure 2.2 (d)) based on the maximum of (potentially different) nonparametric test statistics in each component, which is evaluated using the permutation principle.

- Parametric procedures: Logan (2003) applies the cone order monotonicity principle to the two-sample case with normal observations. Further parametric proposals are mentioned by Tamhane and Logan (2004).

4.3 Multi-Sample Problem

- Dietz (1989) considers a generalization of the Jonckheere–Terpstra test to the multivariate setting. She proposes a sum of coordinatewise Jonckheere

statistics (for the alternative that each variable is stochastically increasing in the given order of groups) and a quadratic form of coordinatewise Jonckheere statistics (for the alternative that each variable is stochastically ordered, possibly in different directions).

- A multivariate Jonckheere–Terpstra test is also proposed by Tsai and Koziol (1994) for the alternative of increasing location parameters in the given order of groups. Their approach is also based on coordinatewise statistics.

- Choi and Marden (1997) propose another generalization of the Jonckheere–Terpstra statistic for the multivariate case. The statistic is obtained by replacing $\operatorname{sgn}(x - y)$ by the spatial sign $(\boldsymbol{x} - \boldsymbol{y})/\left\|\boldsymbol{x} - \boldsymbol{y}\right\|$, i. e. the unit vector pointing from one observation to another (see Definition 7.0.1).

- Tsai and Sen (1990) discuss two tests for randomized block designs with multivariate observations for some rather general alternative. The two tests are based on coordinatewise intra-block rankings and coordinatewise aligned rankings over all blocks, respectively.

- Möttönen, Hüsler, and Oja (2003) present a rotation invariant multivariate analogue of the Page test based on spatial ranks (see Definition 7.0.1).

4.4 Monotone Trend Problem

- Dietz and Killeen (1981) give a test for detecting a monotone trend present in one or more of p variables observed k times. The direction of the trend in each component is not specified in advance. Their statistic is based on rankings within each component and combines within-component statistics by a quadratic form.

4.5 Combination of Univariate Tests

A generally applicable principle to construct multivariate tests is the combination of univariate test results, possibly from different tests for each component.

- One approach is to adjust the univariate p-values in order to keep the overall significance level when multiple tests are performed. Such a procedure only implicitly supplies a global test result. A well-known example is the Bonferroni–Holm procedure (Holm, 1979). Other such procedures are mentioned in Reitmeir and Wassmer (1996). An example of a sequential multiple testing procedure that uses the correlation structure of the individual test statistics can be found in Wei, Lin, and Weissfeld (1989).

- Other suggestions combine the individual p-values into one global test – two famous procedures are those by Tippett (1937) and Fisher (1932), further methods in the field of meta-analysis can be found in Hedges and Olkin (1985). For a recent comparison of several well-known methods, see Loughin (2004).

  However, the usual meta-analysis procedures are not suitable for combining univariate tests into a multivariate test because they are designed for the combination of independent studies, whereas the tests on different variables of one data set generally depend on each other.

  An example of an approach that combines dependent p-values (with an application to order-restricted testing problems) can be found in Kost and McDermott (2002).

We will return to methods for the combination of univariate tests in Chapter 5.

4.6 Related Problems

- The univariate k-sample problem has some connection to the $(k-1)$-variate one-sample problem if the location differences of consecutive samples are considered. One of several examples of tests for this problem can be found in De (1976).

- Parametric procedures for the univariate monotone trend problem (and related problems with a separate location parameter for each observation) can be found in Bartholomew (1961) and Abelson and Tukey (1963).

Chapter 5

Union-Intersection and Intersection-Union Tests

Probably the most obvious way to deal with orthant alternatives is the combination of univariate tests for each component. In this chapter, we present such methods and the more generally applicable union-intersection and intersection-union principles.

5.1 Componentwise Procedures for the $(2^p - 1)$-Orthant Alternative

For the problem of testing

$$\begin{aligned} H_0 &: \boldsymbol{\vartheta} \leq \mathbf{0} \quad \text{vs.} \\ H_1 &: \exists\, i \in \{1, \ldots, p\} : \vartheta_i > 0, \end{aligned}$$

we can use multiple test procedures to combine individual (univariate) tests for

$$\begin{aligned} H_{0i} &: \vartheta_i \leq 0 \quad \text{vs.} \\ H_{1i} &: \vartheta_i > 0. \end{aligned}$$

(Note that open alternative regions are used here, which are more suitable for the multiple comparisons approach.) We then reject the global null hypothesis $H_0 = \bigcap_{i=1}^{p} H_{0i}$ if we can reject at least one of the component null hypotheses (at a suitably corrected level). For both the Bonferroni procedure and the improved version due to Holm (1979), this condition is equivalent to the smallest p-value being less than α/p, where α is the desired level of the (global) test. Pocock, Geller, and Tsiatis (1987) show that even e. g. in the case of a multivariate normal distribution with known correlations, where better decision rules can be worked out, the Bonferroni correction is not overly conservative for moderately correlated components.

We can derive a more general rule from the statements in Rüger (1978): H_0 is rejected if at least k of the p component hypotheses are rejected at the level $k\alpha/p$, where $k \in \{1, \ldots, p\}$ has to be chosen in advance. The choice of k allows for some emphasis on the detection either of small effects in many dimensions or of large effects in few dimensions.

Hommel (1983) modifies Rüger's procedure: For every $k \in \{1, \ldots, p\}$, he compares the k-th smallest individual p-value to $\delta_k := k\alpha/(p \sum_{i=1}^{p} i^{-1})$ and rejects H_0 if any of the ordered p-values is below the corresponding δ_k. This rejection rule allows for a detection of small effects in many dimensions as well as large effects in few dimensions, and in contrast to Rüger's method, no additional parameter has to be chosen.

For a class of rejection rules that includes the Bonferroni, Rüger, and Hommel methods as special cases, see Röhmel and Streitberg (1987).

The method for component hypotheses that are ordered a priori with respect to their importance (Maurer, Hothorn, and Lehmacher, 1995; see Section 5.2) is not appropriate here; its application would consist in a univariate test in the most important component only.

5.2 Componentwise Procedures for the One-Orthant Alternative

Using an open alternative region again, we can formulate the problem as

$$\begin{aligned} H_0 &: \exists\, i \in \{1, \ldots, p\} : \vartheta_i \leq 0 \quad \text{vs.} \\ H_1 &: \boldsymbol{\vartheta} > \mathbf{0}. \end{aligned}$$

Obviously, we can again translate this into a multiple test problem with

$$\begin{aligned} H_{0i} &: \vartheta_i \leq 0 \quad \text{vs.} \\ H_{1i} &: \vartheta_i > 0, \end{aligned}$$

but now the rejection of the original null hypothesis corresponds to the rejection of all partial null hypotheses. Therefore, the Bonferroni, Holm, Rüger, and Hommel approaches generally lead to much too conservative tests.

In fact, such a correction is not necessary at all: Maurer, Hothorn, and Lehmacher (1995) show that the global level α is respected in a multiple comparisons problem if the partial hypotheses are tested sequentially, each at the level α, where the hypotheses are ordered a priori with respect to their importance, and a hypothesis is only tested if all preceding hypotheses have been rejected. (This is a special case of the closed testing procedure described by Marcus, Peritz, and Gabriel, 1976.) In the case of the one-orthant alternative, the rejection of all p component hypotheses is necessary. Therefore, the a priori ordering of the hypotheses is unimportant – each component hypothesis needs to be rejected at

the level α in order to reject the original H_0. (This is also a special case of the method by Rüger, 1978, with $k = p$.) This method is still conservative, i. e. its true level will be strictly less than α except for degenerate examples.

The principle of rejecting the global null hypothesis if and only if each component hypothesis can be rejected is known as the *min test* in the literature (e. g. Sen and Silvapulle, 2002): If the same test statistic (tending to result in larger values for larger ϑ_i's) is calculated for each component, the minimum of them can be compared to the $1 - \alpha$ quantile of the univariate test statistic used.

More efficient tests based on componentwise statistics might be obtained by resampling methods; see e. g. Westfall and Young (1993).

5.3 Union-Intersection Tests

We can apply ideas similar to those presented in Section 5.1 in a more general context: If a null hypothesis can be written as $H_0 = \bigcap_{i \in I} H_{0i}$, we can use the so-called union-intersection (UI) principle to construct a test bases on this representation of H_0. This test rejects H_0 if any of the individual hypotheses H_{0i} can be rejected using a suitable (e. g. univariate) test, i. e. the rejection region of the UI test for H_0 is the union of the individual rejection regions for the H_{0i}'s; see e. g. Mardia, Kent, and Bibby (1979), pp. 129ff. This principle was formulated by Roy (1953).

The level α of the UI test for H_0 is at least the level α^* of the tests for each H_{0i}, but may be much larger than α^*, such that a suitable choice of α^* has to be made. For a finite index set I, the multiple test procedures mentioned in Section 5.1 can be used again.

5.4 Intersection-Union Tests

An analogous procedure to the UI test has been proposed by Berger (1982) and Berger and Sinclair (1984) for the case that $H_0 = \bigcup_{i \in I} H_{0i}$. The rejection region for H_0 of the intersection-union (IU) test is the intersection of the rejection regions of the tests for H_{0i}.

Again, the level α of the IU test is usually not equal to the level α^* of the individual tests. However, α^* is an upper bound for α, and in many cases, α^* cannot be chosen higher than the target level α unless restrictive assumptions, e. g. on the covariance structure, are made. Therefore, $\alpha^* = \alpha$ is often chosen, usually yielding a slightly conservative test.

Tamhane and Logan (2004) combine the UI and the IU approach into a test for non-inferiority in all components of a multivariate parameter and superiority in at least one component.

Chapter 6

Binomial Tests

In this chapter, we discuss several simple one-sample location test statistics that are based on point counts. From these statistics, we can construct tests using binomial null distributions.

General Assumption
In this and the following chapters, unless otherwise indicated, observations will be assumed to come from an absolutely continuous distribution on $\mathbb{R}^p$ that is (at least) directionally symmetric with respect to $\boldsymbol{\vartheta}$.

6.1 Half-Space

Let C be a closed half-space in $\mathbb{R}^p$ including $\mathbf{0}$ in its boundary. When testing $H_0 : \boldsymbol{\vartheta} \in \{\mathbf{0}\} \cup (\mathbb{R}^p \setminus C)$ vs. $H_1 : \boldsymbol{\vartheta} \in C \setminus \{\mathbf{0}\}$, we can obviously count the number of points in C to get a sensible statistic. If n points from a directionally symmetric distribution (with center $\mathbf{0}$) are observed, the probability for each point to fall into C is 0.5, and the number of points in C, $\sum_{i=1}^{n} 1(\boldsymbol{X}_i \in C)$, has a $\mathrm{Bin}(n, 0.5)$ distribution. Therefore, we will reject the null hypothesis if the number of points in C exceeds the $(1-\alpha)$-quantile of the binomial distribution with the given parameters, where α is the desired level of the test.

For every $\boldsymbol{\vartheta} \in \mathbb{R}^p \setminus C$, the probability for a point to fall into C is less than or equal to 0.5. Thus, the specified level of the test will be respected for the whole set constituting the null hypothesis, and not only for $\boldsymbol{\vartheta} = \mathbf{0}$.

This test is affine invariant: If a point $\boldsymbol{x}$ is in C, $T(\boldsymbol{x})$ will also be in $T(C)$ under any affine transformation T.

6.2 Cone/Opposite Cone

If C is no longer a half-space, but some general closed convex cone, the number of points in C still follows a binomial distribution, but the "success probability"

under $\boldsymbol{\vartheta} = \mathbf{0}$, $\mathrm{P}_{\mathbf{0}}(\boldsymbol{X}_i \in C)$, is no longer known. For distributions in the spherical directions class, we could use the proportion of the unit sphere included in the cone, but for more general distributions, this proportion is meaningless.

One possibility is to compare the number $N_1 = \sum_{i=1}^{n} 1(\boldsymbol{X}_i \in C)$ of points in C to the number N_2 of points in the opposite cone $-C = \{\boldsymbol{x} \in \mathbb{R}^p : -\boldsymbol{x} \in C\}$: If $\boldsymbol{\vartheta} = \mathbf{0}$, conditional on the number $M = N_1 + N_2 = \sum_{i=1}^{n} 1(\boldsymbol{X}_i \in C \cup -C)$ of points in the union of both cones, N_1 has a known binomial distribution again, now with parameters M and 0.5. (For the statistic N_1 to be meaningful, we should assume the distribution of the $\boldsymbol{X}_i$'s to be unimodal.)

For the same reason as in the previous section, this test is affine invariant. Further, the test is cone order monotone in each observation (and therefore also in the sample) with respect to C: Let n_1 and n_2 be the observed values of N_1 and N_2, respectively. By addition of a vector from C, points can be moved out of $-C$ and/or into C, but not in the opposite direction, i. e. n_1 can increase by 1 and/or n_2 can decrease by 1. In every possible case, $\mathrm{P}(N_1 \geq n_1 | M = n_1 + n_2)$ remains unchanged or decreases; for the non-trivial cases, this can be seen using Theorem B.2.1.

The obvious drawback of this method is that some of the points are not considered at all, and if the true location parameter lies within $\mathbb{R}^p \setminus (C \cup -C)$, the result will be quite arbitrary. Hence, the test based on this statistic will not be suitable for the (full) composite null hypothesis case, $H_0 : \boldsymbol{\vartheta} \in \{\mathbf{0}\} \cup (\mathbb{R}^p \setminus C)$ vs. $H_1 : \boldsymbol{\vartheta} \in C \setminus \{\mathbf{0}\}$, but only for the simple null hypothesis $H_0 : \boldsymbol{\vartheta} = \mathbf{0}$ or an intermediate composite null hypothesis, such as $H_0 : \boldsymbol{\vartheta} \in -C$.

6.3 Estimated Cone Probability

Another approach to deal with the unknown parameter $\mathrm{P}_{\mathbf{0}}(\boldsymbol{X}_i \in C)$ is to estimate it from the given data. This can be done in the following way:

1. Compute some estimate $\hat{\boldsymbol{\vartheta}}$ for the location parameter, e. g. a multivariate median.

2. As an estimate for $\mathrm{P}_{\mathbf{0}}(\boldsymbol{X}_i \in C)$, take the number of the centered observations $\boldsymbol{X}_i - \hat{\boldsymbol{\vartheta}}$ that are in C, divided by n.

3. Use n and this estimated probability as the binomial parameters to test whether the number of (original) observations $\boldsymbol{X}_i$ in C is significantly too high.

(Modifications: To take advantage from the assumption of directional symmetry, we can also use the combined sample of the centered observations and their reflections and divide by $2n$ in step 2, which leads to more possible values for the estimate of the cone probability and therefore reduces discreteness problems. As

a further alternative, we can use the combined sample of the original observations and their reflections for step 2, such that we can eliminate step 1, which leads to a test resembling the cone/opposite cone test.)

The invariance properties of this method depend on the equivariance properties of the estimate $\hat{\boldsymbol{\vartheta}}$.

Even though this method uses information from all points, it is still not suitable for the general composite null hypothesis case: Imagine the bivariate case with $H_0 : \boldsymbol{\vartheta} \in \{\mathbf{0}\} \cup (\mathbb{R}^2 \setminus C)$ vs. $H_1 : \boldsymbol{\vartheta} \in C \setminus \{\mathbf{0}\}$, where C is the closed first quadrant, and with the uniform distribution on the unit circular disk around $\boldsymbol{\vartheta}$. If the true $\boldsymbol{\vartheta}$ is $(-\varepsilon, 1)^{\mathrm{T}}$, then $\mathrm{P}_{\boldsymbol{\vartheta}}(\boldsymbol{X}_i \in C)$ approaches 0.5 for $\varepsilon \to 0$. But the estimate for $\mathrm{P}_{\mathbf{0}}(\boldsymbol{X}_i \in C)$ will be around 0.25, and so H_0 will be rejected too often, i. e. the specified level for the test will not be respected.

Further, the usage of an estimated cone probability in the test makes it quite difficult to calculate exact rejection probabilities.

Chapter 7

Spatial Sign and Rank Tests

Möttönen and Oja (1995) use the concept of spatial signs and spatial ranks to formulate rotation invariant multivariate analogues of the one- and two-sample sign tests, the Wilcoxon signed rank test, and the Wilcoxon rank sum test for unrestricted alternatives. We will focus on the spatial sign test for the one-sample problem, which we present in Section 7.1. In the following two sections, we propose modifications of this test statistic for one-sided alternatives. At the end of the chapter, we give some remarks concerning analogous modifications of the spatial signed rank test.

We use the following notation throughout this chapter:

Definition 7.0.1. Let $\boldsymbol{x}, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ be p-variate vectors.

(a) The *spatial sign* of $\boldsymbol{x}$ is

$$\operatorname{sgn} \boldsymbol{x} = \begin{cases} \frac{\boldsymbol{x}}{\|\boldsymbol{x}\|} & \|\boldsymbol{x}\| > 0, \\ \mathbf{0} & \|\boldsymbol{x}\| = 0. \end{cases}$$

(b) The *spatial rank* of $\boldsymbol{x}$ with respect to $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ is

$$\operatorname{rnk}(\boldsymbol{x} | \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) = \frac{1}{n} \sum_{i=1}^{n} \operatorname{sgn}(\boldsymbol{x} - \boldsymbol{x}_i). \qquad \triangle$$

The definition for the spatial rank given here is the one used in more recent papers (e. g. Möttönen, Oja, and Tienari, 1997), while in Möttönen and Oja (1995), $-n \operatorname{rnk}(\boldsymbol{x} | \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ is called the spatial rank.

7.1 Spatial Sign Test for an Unrestricted Alternative

Möttönen and Oja (1995) propose a rotation invariant sign test for $H_0 : \boldsymbol{\vartheta} = \mathbf{0}$ vs. $H_1 : \boldsymbol{\vartheta} \neq \mathbf{0}$, where $\boldsymbol{\vartheta}$ is the location parameter of a diagonally symmetric,

absolutely continuous distribution F. Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be a random sample from F. The test is based on the fact that $\sqrt{n}\,\mathrm{rnk}(\boldsymbol{0}|\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)$ has an $\mathrm{N}_p(\boldsymbol{0}, \mathrm{B}_1)$ limiting distribution under the null hypothesis, where $\mathrm{B}_1 = \mathrm{E}_0[\mathrm{sgn}\,\boldsymbol{X}_1(\mathrm{sgn}\,\boldsymbol{X}_1)^{\mathrm{T}}]$. The covariance matrix B_1 can be estimated consistently under H_0 by $\widehat{\mathrm{B}}_{1,n} = n^{-1}\sum_{i=1}^{n}\mathrm{sgn}\,\boldsymbol{X}_i(\mathrm{sgn}\,\boldsymbol{X}_i)^{\mathrm{T}}$, and

$$n\,(\mathrm{rnk}(\boldsymbol{0}|\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n))^{\mathrm{T}}\widehat{\mathrm{B}}_{1,n}^{-1}\,\mathrm{rnk}(\boldsymbol{0}|\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)$$

asymptotically has a chi-square distribution with p degrees of freedom.

7.2 Modification for a Direction Alternative

Geometrically, we can interpret $(\mathrm{rnk}(\boldsymbol{0}|\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n))^{\mathrm{T}}\widehat{\mathrm{B}}_{1,n}^{-1}\,\mathrm{rnk}(\boldsymbol{0}|\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)$ as the L_2 norm of the spatial rank of $\boldsymbol{0}$ after an affine transformation $T : \boldsymbol{x} \mapsto \widehat{\mathrm{B}}_{1,n}^{-1/2}\boldsymbol{x}$. ($\widehat{\mathrm{B}}_{1,n}$ is symmetric and nonnegative definite; if it is positive definite, $\widehat{\mathrm{B}}_{1,n}^{-1}$ exists, is also symmetric and positive definite and has a symmetric square root – see e. g. Harville, 1997, Section 21.9.)

Instead of the norm, we can consider a projection onto a specified direction, which leads us to a test statistic for a restricted alternative: Let a vector $\boldsymbol{v} \neq \boldsymbol{0}$ indicate the assumed direction for the deviation of the location from $\boldsymbol{0}$ under the alternative. The projection that we need is the scalar product of $T(\mathrm{rnk}(\boldsymbol{0}|\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n))$ and a unit vector in the direction of $T(\boldsymbol{v})$:

$$\left(\mathrm{sgn}\left(\widehat{\mathrm{B}}_{1,n}^{-1/2}\boldsymbol{v}\right)\right)^{\mathrm{T}}\widehat{\mathrm{B}}_{1,n}^{-1/2}\,\mathrm{rnk}(\boldsymbol{0}|\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)$$

As a test statistic for the direction alternative, we use

$$S_{\mathrm{sgn,d}}(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n; \boldsymbol{v}) = -\sqrt{n}\left(\mathrm{sgn}\left(\widehat{\mathrm{B}}_{1,n}^{-1/2}\boldsymbol{v}\right)\right)^{\mathrm{T}}\widehat{\mathrm{B}}_{1,n}^{-1/2}\,\mathrm{rnk}(\boldsymbol{0}|\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n),$$

which tends to take positive values if $\boldsymbol{\vartheta}$ lies in the direction of $\boldsymbol{v}$. More precisely, the alternative is $H_1 : \boldsymbol{v}^{\mathrm{T}}[\mathrm{E}(\mathrm{sgn}\,\boldsymbol{X}_i(\mathrm{sgn}\,\boldsymbol{X}_i)^{\mathrm{T}})]^{-1}\tilde{\boldsymbol{\vartheta}} > 0$, where $\tilde{\boldsymbol{\vartheta}}$ is the expected spatial sign of the observations; i. e., under H_1, $\tilde{\boldsymbol{\vartheta}}$ lies within the half-space given by $\boldsymbol{v}$ and the scalar product based on $[\mathrm{E}(\mathrm{sgn}\,\boldsymbol{X}_i(\mathrm{sgn}\,\boldsymbol{X}_i)^{\mathrm{T}})]^{-1}$. Möttönen and Oja (1995) claim that $-\,\mathrm{rnk}(\boldsymbol{0}|\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)$ points "towards the mass of the sample"; however, the exact meaning of this statement is uncertain – at least, examples can be constructed where $\tilde{\boldsymbol{\vartheta}} = \mathrm{E}(\mathrm{sgn}\,\boldsymbol{X}_1) = \mathrm{E}[-\,\mathrm{rnk}(\boldsymbol{0}|\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)]$ and the symmetry point $\boldsymbol{\vartheta}$ of the distribution of $\boldsymbol{X}_1$ are nearly orthogonal to each other. See Appendix A.2 for details.

A related approach (for the bivariate case) is also used by Brown (1983), but the statistic considered there is the sum of scalar products of the (untransformed) spatial signs of the observations and a (untransformed) unit vector in the specified direction.

In order to determine the limiting distribution of $S_{\mathrm{sgn,d}}(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n; \boldsymbol{v})$ for $n \to \infty$ under the simple null hypothesis $H_0 : \tilde{\boldsymbol{\vartheta}} = \boldsymbol{0}$ (which is implied by $\boldsymbol{\vartheta} = \boldsymbol{0}$; see Appendix A.1), we establish a more general lemma first.

Lemma 7.2.1. *Let $(\boldsymbol{X}_n)_{n\in\mathbb{N}}$ be a sequence of random vectors in $\mathbb{R}^p$, $\boldsymbol{X}_n \xrightarrow{\mathrm{d}} \mathrm{N}_p(\boldsymbol{0}, \Sigma)$ $(n \to \infty)$, Σ positive definite, $\Sigma_n \xrightarrow{\mathrm{p}} \Sigma$ $(n \to \infty)$, Σ_n positive definite w. p. 1 for sufficiently large n, $\boldsymbol{v} \in \mathbb{R}^p \setminus \{\boldsymbol{0}\}$ a fixed vector. Then*

$$\left(\operatorname{sgn}\left(\Sigma_n^{-1/2}\boldsymbol{v}\right)\right)^{\mathrm{T}} \Sigma_n^{-1/2}\boldsymbol{X}_n \xrightarrow{\mathrm{d}} \mathrm{N}(0,1) \quad (n \to \infty)$$

(the limit being well defined w. p. 1).

Proof. W. p. 1, there exists $n_0 \in \mathbb{N}$ such that every Σ_n with $n \geq n_0$ is positive definite (and therefore has an inverse Σ_n^{-1}). $\Sigma_n \xrightarrow{\mathrm{p}} \Sigma$ implies $\Sigma_n^{-1} \xrightarrow{\mathrm{p}} \Sigma^{-1}$ by Theorem B.1.2 (b) because $\mathrm{A} \mapsto \mathrm{A}^{-1}$ is continuous for positive definite A. Therefore, by Theorem B.1.4 (c), $\Sigma_n^{-1/2}\boldsymbol{X}_n \xrightarrow{\mathrm{d}} \mathrm{N}_p(\boldsymbol{0}, \mathrm{I}_p)$.

Repeated application of Theorem B.1.2 (a) yields

$$\operatorname{sgn}\left(\Sigma_n^{-1/2}\boldsymbol{v}\right) \xrightarrow{\mathrm{d}} \operatorname{sgn}\left(\Sigma^{-1/2}\boldsymbol{v}\right).$$

Because $\operatorname{sgn}\left(\Sigma^{-1/2}\boldsymbol{v}\right)$ is constant, we can apply Theorem B.1.1 in order to obtain

$$\left(\Sigma_n^{-1/2}\boldsymbol{X}_n, \operatorname{sgn}\left(\Sigma_n^{-1/2}\boldsymbol{v}\right)\right) \xrightarrow{\mathrm{d}} \left(\boldsymbol{Z}_p, \operatorname{sgn}\left(\Sigma^{-1/2}\boldsymbol{v}\right)\right),$$

where $\boldsymbol{Z}_p \sim \mathrm{N}_p(\boldsymbol{0}, \mathrm{I}_p)$.

Finally, Theorem B.1.2 (a) yields

$$\left(\operatorname{sgn}\left(\Sigma_n^{-1/2}\boldsymbol{v}\right)\right)^{\mathrm{T}} \Sigma_n^{-1/2}\boldsymbol{X}_n \xrightarrow{\mathrm{d}} \left(\operatorname{sgn}\left(\Sigma^{-1/2}\boldsymbol{v}\right)\right)^{\mathrm{T}} \boldsymbol{Z}_p.$$

But the projection of an $\mathrm{N}_p(\boldsymbol{0}, \mathrm{I}_p)$ random vector onto any direction is $\mathrm{N}(0,1)$. □

For $\boldsymbol{v} = (1, \ldots, 1)^{\mathrm{T}}$, O'Brien (1984) as well as Pocock, Geller, and Tsiatis (1987) use this type of asymptotically standard normal statistics.

Theorem 7.2.2. *Under H_0, there exists w. p. 1 a number $n_0 \in \mathbb{N}$ such that $S_{\mathrm{sgn,d}}(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n; \boldsymbol{v})$ is well defined for all $n \geq n_0$, and it converges in distribution to standard normal for $n \to \infty$.*

Proof. It has already been mentioned that the limiting distribution of $\sqrt{n}\,\mathrm{rnk}(\boldsymbol{0}|\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)$ is $\mathrm{N}_p(\boldsymbol{0}, \mathrm{B}_1)$ (central limit theorem). The distribution of the $\boldsymbol{X}_i$'s is absolutely continuous and therefore genuinely p-variate, and so is the distribution of the $\operatorname{sgn}\boldsymbol{X}_i$'s. Therefore, B_1 is positive definite. Further, $\widehat{\mathrm{B}}_{1,n} = n^{-1}\sum_{i=1}^n \operatorname{sgn}\boldsymbol{X}_i(\operatorname{sgn}\boldsymbol{X}_i)^{\mathrm{T}}$ is a consistent estimate for B_1 under H_0 and positive definite w. p. 1 for $n > p$. Therefore, Lemma 7.2.1 is applicable, which completes the proof. □

The directional version of the spatial sign test shares the invariance property of the unrestricted version by Möttönen and Oja (1995):

Theorem 7.2.3. *$S_{\mathrm{sgn,d}}(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n; \boldsymbol{v})$ is orthogonally invariant.*

Proof. Let A be an orthogonal $p \times p$ matrix, i. e. $\mathrm{A}^{\mathrm{T}}\mathrm{A} = \mathrm{A}\mathrm{A}^{\mathrm{T}} = \mathrm{I}_p$. Recalling the respective definitions, we can easily see that

$$\begin{aligned} \operatorname{sgn}(\mathrm{A}\boldsymbol{x}) &= \mathrm{A} \operatorname{sgn} \boldsymbol{x}, \\ \operatorname{rnk}(\mathbf{0}|\mathrm{A}\boldsymbol{x}_1, \ldots, \mathrm{A}\boldsymbol{x}_n) &= \mathrm{A} \operatorname{rnk}(\mathbf{0}|\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n), \\ \widehat{\mathrm{B}}^*_{1,n} &= \mathrm{A}\widehat{\mathrm{B}}_{1,n}\mathrm{A}^{\mathrm{T}}, \end{aligned}$$

where $\widehat{\mathrm{B}}^*_{1,n} = n^{-1} \sum_{i=1}^n \operatorname{sgn}(\mathrm{A}\boldsymbol{X}_i)(\operatorname{sgn}(\mathrm{A}\boldsymbol{X}_i))^{\mathrm{T}}$. Further,

$$\widehat{\mathrm{B}}^{*\,-1/2}_{1,n} = \mathrm{A}\widehat{\mathrm{B}}^{-1/2}_{1,n}\mathrm{A}^{\mathrm{T}},$$

and, by combination of these equalities,

$$\begin{aligned} &S_{\mathrm{sgn,d}}(\mathrm{A}\boldsymbol{X}_1, \ldots, \mathrm{A}\boldsymbol{X}_n; \mathrm{A}\boldsymbol{v}) \\ &\quad = -\sqrt{n} \left(\operatorname{sgn} \left(\widehat{\mathrm{B}}^{*\,-1/2}_{1,n} \mathrm{A}\boldsymbol{v} \right) \right)^{\mathrm{T}} \widehat{\mathrm{B}}^{*\,-1/2}_{1,n} \operatorname{rnk}(\mathbf{0}|\mathrm{A}\boldsymbol{X}_1, \ldots, \mathrm{A}\boldsymbol{X}_n) \\ &\quad = -\sqrt{n} \left(\operatorname{sgn} \left(\mathrm{A}\widehat{\mathrm{B}}^{-1/2}_{1,n} \mathrm{A}^{\mathrm{T}}\mathrm{A}\boldsymbol{v} \right) \right)^{\mathrm{T}} \mathrm{A}\widehat{\mathrm{B}}^{-1/2}_{1,n} \mathrm{A}^{\mathrm{T}}\mathrm{A} \operatorname{rnk}(\mathbf{0}|\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n) \\ &\quad = -\sqrt{n} \left(\mathrm{A} \operatorname{sgn} \left(\widehat{\mathrm{B}}^{-1/2}_{1,n} \boldsymbol{v} \right) \right)^{\mathrm{T}} \mathrm{A}\widehat{\mathrm{B}}^{-1/2}_{1,n} \operatorname{rnk}(\mathbf{0}|\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n) \\ &\quad = -\sqrt{n} \left(\operatorname{sgn} \left(\widehat{\mathrm{B}}^{-1/2}_{1,n} \boldsymbol{v} \right) \right)^{\mathrm{T}} \mathrm{A}^{\mathrm{T}}\mathrm{A}\widehat{\mathrm{B}}^{-1/2}_{1,n} \operatorname{rnk}(\mathbf{0}|\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n) \\ &\quad = S_{\mathrm{sgn,d}}(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n; \boldsymbol{v}), \end{aligned}$$

which is the desired invariance property. □

7.3 Adaptation to a Sector Alternative

Cone alternatives are more important than direction alternatives in many applications, such that it would be desirable to extend the spatial sign test to this more general problem. As this is the simplest case, we only treat sector alternatives for the bivariate case here.

A convex sector can be described as the intersection of two half-planes. Therefore, again with $\tilde{\boldsymbol{\vartheta}} = \mathrm{E}(\operatorname{sgn} \boldsymbol{X}_1)$, we construct a test for the simple $H_0 : \tilde{\boldsymbol{\vartheta}} = \mathbf{0}$ vs. $H_1 : \tilde{\boldsymbol{\vartheta}} \in \Theta_1$ (Θ_1 being a sector with an angle strictly less than π) from tests for the alternative that the expected spatial sign lies within some half-plane. (Such half-plane alternatives are related to direction alternatives, but it should be noted that the directional version of the spatial sign test described above does not test

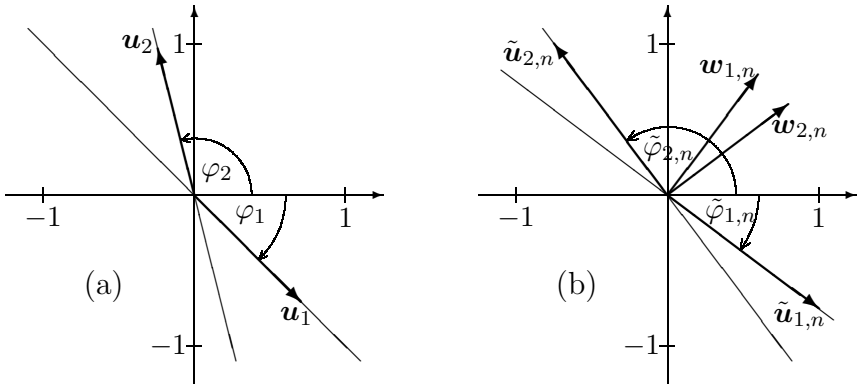

Figure 7.1: (a) Sector alternative in the original coordinate system; (b) Transformed sector alternative: $\tilde{\boldsymbol{u}}_{i,n} = \widehat{\mathrm{B}}_{1,n}^{-1/2}\boldsymbol{u}_i$.

whether the expected spatial sign lies within an explicitly given half-plane; instead, the hypothetical direction $\boldsymbol{v}$ is given, which is transformed using $\widehat{\mathrm{B}}_{1,n}^{-1/2}$, and in this transformed space, a half-plane can be associated with the direction $\widehat{\mathrm{B}}_{1,n}^{-1/2}\boldsymbol{v}$. This does not lead to the same result as if we transformed the half-space given by $\boldsymbol{v}$.)

We use the following notation (cf. Figure 7.1): The sector Θ_1 is given as the area between the two angles $\varphi_1 < \varphi_2$, corresponding to unit vectors $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$. These vectors are transformed to $\tilde{\boldsymbol{u}}_{i,n} = \widehat{\mathrm{B}}_{1,n}^{-1/2}\boldsymbol{u}_i$ (not necessarily of length 1; with angles $\tilde{\varphi}_{i,n}$), such that $\tilde{\boldsymbol{u}}_{1,n}$ and $\tilde{\boldsymbol{u}}_{2,n}$ indicate the borders of the transformed sector. Unit vectors $\boldsymbol{w}_{1,n}$ and $\boldsymbol{w}_{2,n}$ are defined such that they are orthogonal on the borders and pointing towards the transformed sector, i. e. $\boldsymbol{w}_{i,n}^{\mathrm{T}}\tilde{\boldsymbol{u}}_{i,n} = 0$ and $\boldsymbol{w}_{i,n}^{\mathrm{T}}\tilde{\boldsymbol{u}}_{j,n} > 0$ for $i \neq j$.

If $-\sqrt{n}\widehat{\mathrm{B}}_{1,n}^{-1/2}\,\mathrm{rnk}(\mathbf{0}|\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n)$ lies within the transformed sector, we consider this as an indication that the location parameter $\boldsymbol{\vartheta}$ lies within the given sector. Hence, both projections (on the directions given by $\boldsymbol{w}_{1,n}$ and $\boldsymbol{w}_{2,n}$) of this transformed spatial rank should be (substantially) greater than 0 in order to reject H_0. The minimum of these two projections appears to be a sensible test statistic:

$$S_{\mathrm{sgn,s}}(\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n;\varphi_1,\varphi_2) = \min\Big(-\sqrt{n}\boldsymbol{w}_{1,n}^{\mathrm{T}}\widehat{\mathrm{B}}_{1,n}^{-1/2}\,\mathrm{rnk}(\mathbf{0}|\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n), \\ -\sqrt{n}\boldsymbol{w}_{2,n}^{\mathrm{T}}\widehat{\mathrm{B}}_{1,n}^{-1/2}\,\mathrm{rnk}(\mathbf{0}|\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n)\Big)$$

Even though we are looking at a simple H_0, the use of this minimum of two test statistics can be interpreted as an intersection-union test.

Theorem 7.3.1. *Let $\varphi_2 - \varphi_1 < \pi$, $\boldsymbol{w}_{1,n}, \boldsymbol{w}_{2,n}$ be defined as above, and let $\boldsymbol{w}_1, \boldsymbol{w}_2$ be defined analogously, but based on B_1 instead of $\widehat{\mathrm{B}}_{1,n}$.*

Under H_0, there exists w. p. 1 an $n_0 \in \mathbb{N}$ such that $S_{\text{sgn,s}}(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n; \varphi_1, \varphi_2)$ is well defined for all $n \geq n_0$, and its limiting density function for $n \to \infty$ is

$$f_{S_{\text{sgn,s}}}(y) = 2\phi(y) \left[1 - \Phi\left(y\sqrt{\frac{1-\rho}{1+\rho}}\right)\right],$$

where $\rho = \boldsymbol{w}_1^{\mathrm{T}}\boldsymbol{w}_2$ and ϕ, Φ are the standard normal density function and cumulative distribution function, respectively.

Proof. For the existence w. p. 1 of n_0 with the stated property, see the proof of Theorem 7.2.2. From the proofs of Lemma 7.2.1 and Theorem 7.2.2, we know that $-\sqrt{n}\widehat{\mathrm{B}}_{1,n}^{-1/2}\,\mathrm{rnk}(\mathbf{0}|\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)$ has an $\mathrm{N}_2(\mathbf{0}, \mathrm{I}_2)$ limiting distribution. $\widehat{\mathrm{B}}_{1,n} \xrightarrow{\mathrm{p}} \mathrm{B}_1$ implies, by Theorem B.1.2 (a), that $\boldsymbol{w}_{i,n} \xrightarrow{\mathrm{d}} \boldsymbol{w}_i$. Theorem B.1.1 ensures that

$$\left(-\sqrt{n}\widehat{\mathrm{B}}_{1,n}^{-1/2}\,\mathrm{rnk}(\mathbf{0}|\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n), \boldsymbol{w}_{1,n}, \boldsymbol{w}_{2,n}\right) \xrightarrow{\mathrm{d}} (\boldsymbol{Z}_2, \boldsymbol{w}_1, \boldsymbol{w}_2)$$

with $\boldsymbol{Z}_2 \sim \mathrm{N}_2(\mathbf{0}, \mathrm{I}_2)$, and according to Theorem B.1.2 (a),

$$S_{\text{sgn,s}}(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n; \varphi_1, \varphi_2) \xrightarrow{\mathrm{d}} \min\left(\boldsymbol{w}_1^{\mathrm{T}}\boldsymbol{Z}_2, \boldsymbol{w}_2^{\mathrm{T}}\boldsymbol{Z}_2\right).$$

Theorem B.3.3 yields the desired statement (note that $\boldsymbol{w}_1^{\mathrm{T}}\boldsymbol{w}_2$ is the cosine of the angle of the transformed sector, $\tilde{\varphi}_2 - \tilde{\varphi}_1$). □

Theorem 7.3.2. *$S_{\text{sgn,s}}(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n; \varphi_1, \varphi_2)$ is orthogonally invariant.*

Proof. The proof of rotation invariance is analogous to the one of Theorem 7.2.3. For the invariance under some reflection, e. g. $(x_1, x_2)^{\mathrm{T}} \mapsto (x_1, -x_2)^{\mathrm{T}}$, the proof is similar, but we have to relabel the angles specifying the sector in order to maintain the condition $\varphi_1 < \varphi_2$. As soon as invariance under every rotation and under some reflection is assured, the proof of orthogonal invariance is complete. □

Theorem 7.3.3. *Under H_0, $S_{\text{sgn,s}}(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n; \varphi_1, \varphi_2)$ is distribution-free for distributions of the $\boldsymbol{X}_i$'s belonging to the spherical directions class.*

Proof. For $\boldsymbol{X}_i$'s coming from a distribution with spherical directions, $\operatorname{sgn}\boldsymbol{X}_i$ is uniformly distributed on the unit circle. Since the influence of $\boldsymbol{X}_i$ on $S_{\text{sgn,s}}$ is only through $\operatorname{sgn}\boldsymbol{X}_i$, the statement follows immediately. □

According to Theorem 7.3.3, critical values for a test based on $S_{\text{sgn,s}}$ could be tabulated for the class of distributions with spherical directions, depending only on the sample size and the angle of the sector. For the much more general class of directionally symmetric distributions, a permutation test could also be used to deal with small sample situations: Multiplication of an arbitrary subset of observations by -1 results in the multiplication by -1 of the corresponding spatial signs, but does not affect $\widehat{\mathrm{B}}_{1,n}$, such that the most complex parts of the calculation

need only be done once. (Both Theorem 7.3.3 and the remark concerning the permutation test apply to $S_{\text{sgn,d}}$ as well.)

It is not difficult to find examples showing that $S_{\text{sgn,s}}(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n; \varphi_1, \varphi_2)$ and the tests based on it are not cone order monotone in the sample with respect to Θ_1 or Θ_1^* (and therefore not cone order monotone in each observation either).

Tests based on $S_{\text{sgn,s}}(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n; \varphi_1, \varphi_2)$ are suitable for convex sectors only. However, for non-convex sectors (i. e. the case $\varphi_2 - \varphi_1 > \pi$), we can construct a similar statistic using the maximum of two projections instead of the minimum, i. e. based on the union-intersection principle.

Both tests (based on the minimum or the maximum of projections) can be generalized to the multivariate case, where the alternative regions are convex polyhedral cones or their complements, respectively. While the exact distribution of the test statistic will be difficult to derive, the permutation test approach outlined above can be used.

7.4 Spatial Signed Rank Test

In a completely analogous way, we can adapt the spatial signed rank test by Möttönen and Oja (1995) to one-sided alternatives.

For the unrestricted versions, Möttönen, Oja, and Tienari (1997) give the asymptotic efficiencies of spatial sign and signed rank tests. As in the univariate case, the signed rank test is found to be more efficient for normal or moderately heavy-tailed data in low dimensions, but the difference diminishes as the dimension increases, and for heavy-tailed, high-dimensional data, the sign test is even more efficient than the signed rank test.

Similar properties can be expected for the directional versions. Because of the substantially higher computational complexity of the signed rank methods, we do not consider their adaptation to restricted alternatives in detail here.

Chapter 8

Tests Based on the Procedure by Randles (2000)

8.1 Sign Test for an Unrestricted Alternative

Randles (2000) describes an affine invariant multivariate sign test for $H_0 : \boldsymbol{\vartheta} = \mathbf{0}$ vs. $H_1 : \boldsymbol{\vartheta} \neq \mathbf{0}$, $\boldsymbol{\vartheta} \in \mathbb{R}^p$ being the location parameter of a directionally symmetric, absolutely continuous distribution F. He uses a transformation $\boldsymbol{x} \mapsto \widehat{\mathrm{A}}_{\mathrm{d}}\boldsymbol{x}$, where $\widehat{\mathrm{A}}_{\mathrm{d}}$ is the upper triangular $p \times p$ matrix with positive diagonal elements, with a 1 as the upper-left element, and such that

$$\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\widehat{\mathrm{A}}_{\mathrm{d}}\boldsymbol{X}_i}{\|\widehat{\mathrm{A}}_{\mathrm{d}}\boldsymbol{X}_i\|}\right)\left(\frac{\widehat{\mathrm{A}}_{\mathrm{d}}\boldsymbol{X}_i}{\|\widehat{\mathrm{A}}_{\mathrm{d}}\boldsymbol{X}_i\|}\right)^{\mathrm{T}} = \frac{1}{p}\,\mathrm{I}_p.$$

The matrix $\widehat{\mathrm{A}}_{\mathrm{d}}$ is based on a "most robust" estimate of scatter for elliptically symmetric distributions proposed by Tyler (1987).

With $\boldsymbol{V}_i = \widehat{\mathrm{A}}_{\mathrm{d}}\boldsymbol{X}_i/\|\widehat{\mathrm{A}}_{\mathrm{d}}\boldsymbol{X}_i\|$ and $\bar{\boldsymbol{V}} = n^{-1}\sum_{i=1}^{n}\boldsymbol{V}_i$, Randles uses the test statistic

$$Q_{\mathrm{d}} = np\bar{\boldsymbol{V}}^{\mathrm{T}}\bar{\boldsymbol{V}}.$$

Asymptotically for $n \to \infty$, Q_{d} has a chi-square distribution with p degrees of freedom if $\boldsymbol{\vartheta} = \mathbf{0}$. The computational complexity of a conditional test for small samples is reduced by the fact that $\widehat{\mathrm{A}}_{\mathrm{d}}$ is invariant under sign changes of any subset of the observations $\boldsymbol{X}_i$. (Note that the "d" in Q_{d} comes from Randles's paper and indicates the fact that only the directions of the observations from the origin are taken into account, while we normally use a "d" subscript to denote a test for a direction alternative.)

In the spatial sign and rank notation from the preceding chapter, the condition on $\widehat{\mathrm{A}}_{\mathrm{d}}$ can also be written as

$$\frac{1}{n}\sum_{i=1}^{n}\operatorname{sgn}\left(\widehat{\mathrm{A}}_{\mathrm{d}}\boldsymbol{X}_i\right)\left(\operatorname{sgn}\left(\widehat{\mathrm{A}}_{\mathrm{d}}\boldsymbol{X}_i\right)\right)^{\mathrm{T}} = \frac{1}{p}\,\mathrm{I}_p,$$

and the test statistic is

$$Q_{\mathrm{d}} = np \left(\operatorname{rnk}\left(\mathbf{0} \,\middle|\, \widehat{\mathrm{A}}_{\mathrm{d}}\boldsymbol{X}_1, \ldots, \widehat{\mathrm{A}}_{\mathrm{d}}\boldsymbol{X}_n \right) \right)^{\mathrm{T}} \operatorname{rnk}\left(\mathbf{0} \,\middle|\, \widehat{\mathrm{A}}_{\mathrm{d}}\boldsymbol{X}_1, \ldots, \widehat{\mathrm{A}}_{\mathrm{d}}\boldsymbol{X}_n \right).$$

The essential difference to the spatial sign test by Möttönen and Oja (1995) from Section 7.1 is that Randles first standardizes the observations, then calculates the spatial rank of $\mathbf{0}$ with respect to these transformed observations and takes its norm, whereas Möttönen and Oja first calculate the spatial rank of $\mathbf{0}$, then standardize it and take the norm. This reversed order of steps is the reason for the different invariance properties.

8.2 Modification for a Direction Alternative

Like the spatial sign test in Chapter 7, Randles's test statistic involves an affine transformation. Therefore, for a directional version, we also need to transform the hypothetical direction. Once again, we replace the quadratic form by a scalar product:

$$S_{\mathrm{Randles,d}}(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n; \boldsymbol{v}) = \sqrt{np} \left(\frac{\widehat{\mathrm{A}}_{\mathrm{d}}\boldsymbol{v}}{\|\widehat{\mathrm{A}}_{\mathrm{d}}\boldsymbol{v}\|} \right)^{\mathrm{T}} \bar{\boldsymbol{V}},$$

where $\bar{\boldsymbol{V}}$ is the mean of the transformed observations $\boldsymbol{V}_i = \widehat{\mathrm{A}}_{\mathrm{d}}\boldsymbol{X}_i / \|\widehat{\mathrm{A}}_{\mathrm{d}}\boldsymbol{X}_i\|$.

In spatial sign and rank notation:

$$S_{\mathrm{Randles,d}}(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n; \boldsymbol{v}) = -\sqrt{np} \left(\operatorname{sgn}\left(\widehat{\mathrm{A}}_{\mathrm{d}}\boldsymbol{v} \right) \right)^{\mathrm{T}} \operatorname{rnk}\left(\mathbf{0} \,\middle|\, \widehat{\mathrm{A}}_{\mathrm{d}}\boldsymbol{X}_1, \ldots \widehat{\mathrm{A}}_{\mathrm{d}}\boldsymbol{X}_n \right).$$

As for the test statistic for the unrestricted alternative, the major difference to $S_{\mathrm{sgn,d}}(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n; \boldsymbol{v})$ from the preceding chapter is the reversed order of the standardization and the application of the spatial rank. While we have used the square root of a covariance matrix estimate (for the spatial signs) for $S_{\mathrm{sgn,d}}(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n; \boldsymbol{v})$, we maintain $\widehat{\mathrm{A}}_{\mathrm{d}}$ from Randles's proposal here, which is (up to a constant) the Cholesky factorization of a scatter matrix estimate (for the original data). We could also use the square root of this estimate instead.

It is clear that we cannot specify the hypotheses being tested in terms of the Euclidean scalar product in the original data space (i. e. $H_1 : \boldsymbol{v}^{\mathrm{T}}\boldsymbol{\vartheta} > 0$) when we desire affine invariance of a test – orthogonality of directions is not preserved under affine transformations. We therefore have to use some data-based "orthogonality criterion". The alternative that we test here is of the form $H_1 : (\mathrm{A}_0\boldsymbol{v})^{\mathrm{T}} \operatorname{E}[\operatorname{sgn}(\mathrm{A}_0\boldsymbol{X}_1)] > 0$, where A_0 is the population analogue to $\widehat{\mathrm{A}}_{\mathrm{d}}$, i. e. the upper triangular matrix with positive diagonal elements, a 1 as the upper-left element, and such that

$$\operatorname{E}\left[\left(\frac{\mathrm{A}_0\boldsymbol{X}}{\|\mathrm{A}_0\boldsymbol{X}\|} \right) \left(\frac{\mathrm{A}_0\boldsymbol{X}}{\|\mathrm{A}_0\boldsymbol{X}\|} \right)^{\mathrm{T}} \right] = \frac{1}{p}\, \mathrm{I}_p.$$

The corresponding simple null hypothesis is $H_0 : \mathrm{E}[\mathrm{sgn}(\mathrm{A}_0\boldsymbol{X}_1)] = \boldsymbol{0}$, which is implied by $\boldsymbol{\vartheta} = \boldsymbol{0}$ (as in Section 7.2).

Note that with $h(\boldsymbol{x}, \boldsymbol{y}) = [\mathrm{E}(\mathrm{sgn}(\mathrm{A}_0\boldsymbol{x}))]^{\mathrm{T}}\,\mathrm{E}(\mathrm{sgn}(\mathrm{A}_0\boldsymbol{y}))$, we can write H_1 as $h(\boldsymbol{v}, \boldsymbol{X}_1) > 0$. This definition of h enforces symmetry, but h is neither bilinear nor positive definite and therefore not a scalar product (in $\boldsymbol{x}$ and $\boldsymbol{y}$). However, we can interpret h as the Euclidean scalar product in the space of expected spatial signs of the transformed data.

Theorem 8.2.1. *$S_{\mathrm{Randles,d}}(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n; \boldsymbol{v})$ is affine invariant for $n > p(p-1)$.*

Proof. Let D be a nonsingular $p \times p$ matrix, $\boldsymbol{Y}_i = \mathrm{D}\boldsymbol{X}_i$, and let $\widehat{\mathrm{A}}_{\mathrm{d},\boldsymbol{X}}$ and $\widehat{\mathrm{A}}_{\mathrm{d},\boldsymbol{Y}}$ denote the scatter matrices estimated from the $\boldsymbol{X}_i$'s and $\boldsymbol{Y}_i$'s, respectively. Analogously, we use subscripts $\boldsymbol{X}$ and $\boldsymbol{Y}$ to indicate the observations that underlie the calculation of $\boldsymbol{V}_i$ and $\bar{\boldsymbol{V}}$.

Randles shows that the matrix $\Theta = c_0^{-1/2}\widehat{\mathrm{A}}_{\mathrm{d},\boldsymbol{Y}}\mathrm{D}\widehat{\mathrm{A}}_{\mathrm{d},\boldsymbol{X}}^{-1}$ (where $c_0 > 0$ depends on D and the observations $\boldsymbol{X}_i$) is orthogonal for $n > p(p-1)$, and that $\boldsymbol{V}_{i,\boldsymbol{Y}} = \Theta\boldsymbol{V}_{i,\boldsymbol{X}}$. Therefore (using the definition of Θ for the second equality),

$$
\begin{aligned}
S_{\mathrm{Randles,d}}(\mathrm{D}\boldsymbol{X}_1, \ldots, \mathrm{D}\boldsymbol{X}_n; \mathrm{D}\boldsymbol{v}) &= \sqrt{np}\left(\frac{\widehat{\mathrm{A}}_{\mathrm{d},\boldsymbol{Y}}\mathrm{D}\boldsymbol{v}}{\|\widehat{\mathrm{A}}_{\mathrm{d},\boldsymbol{Y}}\mathrm{D}\boldsymbol{v}\|}\right)^{\mathrm{T}}\bar{\boldsymbol{V}}_{\boldsymbol{Y}} \\
&= \sqrt{np}\left(\frac{c_0^{1/2}\Theta\widehat{\mathrm{A}}_{\mathrm{d},\boldsymbol{X}}\boldsymbol{v}}{\|c_0^{1/2}\Theta\widehat{\mathrm{A}}_{\mathrm{d},\boldsymbol{X}}\boldsymbol{v}\|}\right)^{\mathrm{T}}\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{V}_{i,\boldsymbol{Y}} \\
&= \sqrt{np}\left(\frac{\Theta\widehat{\mathrm{A}}_{\mathrm{d},\boldsymbol{X}}\boldsymbol{v}}{\|\widehat{\mathrm{A}}_{\mathrm{d},\boldsymbol{X}}\boldsymbol{v}\|}\right)^{\mathrm{T}}\frac{1}{n}\sum_{i=1}^{n}\Theta\boldsymbol{V}_{i,\boldsymbol{X}} \\
&= \sqrt{np}\left(\frac{\widehat{\mathrm{A}}_{\mathrm{d},\boldsymbol{X}}\boldsymbol{v}}{\|\widehat{\mathrm{A}}_{\mathrm{d},\boldsymbol{X}}\boldsymbol{v}\|}\right)^{\mathrm{T}}\Theta^{\mathrm{T}}\Theta\bar{\boldsymbol{V}}_{\boldsymbol{X}} \\
&= S_{\mathrm{Randles,d}}(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n; \boldsymbol{v}).
\end{aligned}
$$

Because the test for a null location parameter $\boldsymbol{\vartheta}_0 \neq \boldsymbol{0}$ is defined by testing for a null value of $\boldsymbol{0}$ based on the observations $\boldsymbol{X}_i - \boldsymbol{\vartheta}_0$ (see Section 2.4), invariance under translations is obvious, which completes the proof of affine invariance. □

Theorem 8.2.2. *Under H_0 and for a directionally symmetric distribution of the $\boldsymbol{X}_i$'s, the limiting distribution of $S_{\mathrm{Randles,d}}(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n; \boldsymbol{v})$ for $n \to \infty$ is standard normal.*

Proof. We can take the cumbersome part of the proof from that of Theorem 1 in Randles (2000): Define $\boldsymbol{U}_i = \mathrm{A}_0\boldsymbol{X}_i/\|\mathrm{A}_0\boldsymbol{X}_i\|$. According to Randles, $\mathrm{E}[\boldsymbol{U}_i] = \boldsymbol{0}$, $\sqrt{n}(\widehat{\mathrm{A}}_{\mathrm{d}} - \mathrm{A}_0) = \mathrm{O}_p(1)$ $(n \to \infty)$, and therefore, by Lemma A.3 from Randles (2000), $\sqrt{n}(\bar{\boldsymbol{V}} - \bar{\boldsymbol{U}}) \xrightarrow{\mathrm{p}} \boldsymbol{0}$.

Further, the central limit theorem yields $\sqrt{n}\bar{\boldsymbol{U}} \xrightarrow{\mathrm{d}} \mathrm{N}_p(\boldsymbol{0}, \mathrm{E}[\boldsymbol{U}_i\boldsymbol{U}_i^{\mathrm{T}}])$, which is $\mathrm{N}_p(\boldsymbol{0}, \frac{1}{p}\mathrm{I}_p)$ by the definition of A_0. Combining this with Randles's result from above, we obtain $\sqrt{np}\bar{\boldsymbol{V}} \xrightarrow{\mathrm{d}} \mathrm{N}_p(\boldsymbol{0}, \mathrm{I}_p)$ by Theorem B.1.3.

By complete analogy to the proof of Lemma 7.2.1, $S_{\mathrm{Randles,d}}(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n; \boldsymbol{v}) = (\mathrm{sgn}(\widehat{\mathrm{A}}_{\mathrm{d}}\boldsymbol{v}))^{\mathrm{T}}\sqrt{np}\bar{\boldsymbol{V}} \xrightarrow{\mathrm{d}} (\mathrm{sgn}(\mathrm{A}_0\boldsymbol{v}))^{\mathrm{T}}\boldsymbol{Z}_p$, with $\boldsymbol{Z}_p \sim \mathrm{N}_p(\boldsymbol{0}, \mathrm{I}_p)$. This projection has an $\mathrm{N}(0,1)$ distribution. □

8.3 Adaptation to a Sector Alternative

For $p = 2$, as in the case of the spatial sign test by Möttönen and Oja, we can obtain a version for a convex sector alternative again by the consideration of the minimum of two projections of $\sqrt{np}\bar{\boldsymbol{V}} = \sqrt{2n}\bar{\boldsymbol{V}}$. The simple null hypothesis is still $H_0 : \mathrm{E}[\mathrm{sgn}(\mathrm{A}_0\boldsymbol{X}_1)] = \boldsymbol{0}$, and the alternative is now $H_1 : \mathrm{E}[\mathrm{sgn}(\mathrm{A}_0\boldsymbol{X}_1)] \in \mathrm{A}_0\Theta_1$. As above, this is in terms of the expected spatial sign of the transformed observations, now conjectured to lie in the transformed convex sector. An equivalent formulation of the alternative is

$$H_1 : \mathrm{E}\left[\frac{\boldsymbol{X}_1}{\|\mathrm{A}_0\boldsymbol{X}_1\|}\right] = \mathrm{E}\left[\mathrm{sgn}\,\boldsymbol{X}_1\frac{\|\boldsymbol{X}_1\|}{\|\mathrm{A}_0\boldsymbol{X}_1\|}\right] \in \Theta_1,$$

which refers to the spatial signs, but with a weight reflecting the affine transformation necessary to standardize the covariance structure of the spatial signs. The effect of this factor is that more probable directions receive more weight (as long as the underlying distribution is in the elliptical directions class).

We use a notation similar to the one in Section 7.3: The sector Θ_1 (with an angle strictly less than π) is given by the angles φ_1 and φ_2, corresponding to unit vectors $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$. The transformation matrix is now $\widehat{\mathrm{A}}_{\mathrm{d}}$ (based on a sample of size n), and therefore the direction vectors are transformed to $\tilde{\boldsymbol{u}}_{i,n} = \widehat{\mathrm{A}}_{\mathrm{d}}\boldsymbol{u}_i$. Again, we define unit vectors $\boldsymbol{w}_{i,n}$ that are orthogonal on $\tilde{\boldsymbol{u}}_{i,n}$ and pointing towards the sector. The theoretical analogues are $\boldsymbol{w}_i$, defined with A_0 instead of $\widehat{\mathrm{A}}_{\mathrm{d}}$.

We can now define the test statistic as

$$S_{\mathrm{Randles,s}}(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n; \varphi_1, \varphi_2) = \min\left(-\sqrt{2n}\boldsymbol{w}_{1,n}^{\mathrm{T}}\bar{\boldsymbol{V}}, -\sqrt{2n}\boldsymbol{w}_{2,n}^{\mathrm{T}}\bar{\boldsymbol{V}}\right).$$

Theorem 8.3.1. *With the notation from above, assuming $\varphi_2 - \varphi_1 < \pi$, under H_0 and for a directionally symmetric distribution of the $\boldsymbol{X}_i$'s, the limiting density function of $S_{\mathrm{Randles,s}}(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n; \varphi_1, \varphi_2)$ for $n \to \infty$ is*

$$f_{S_{\mathrm{Randles,s}}}(y) = 2\phi(y)\left[1 - \Phi\left(y\sqrt{\frac{1-\rho}{1+\rho}}\right)\right],$$

where $\rho = \boldsymbol{w}_1^{\mathrm{T}}\boldsymbol{w}_2$ and ϕ, Φ are the standard normal density function and cumulative distribution function, respectively.

Proof. Note that $\sqrt{2n}\bar{\boldsymbol{V}}$ has an $\mathrm{N}_2(\mathbf{0}, \mathrm{I}_2)$ limiting distribution and that $\widehat{\mathrm{A}}_{\mathrm{d}} \xrightarrow{\mathrm{p}} \mathrm{A}_0$ $(n \to \infty)$ – see the proof of Theorem 8.2.2. Thus, we can complete the proof analogously to the one of Theorem 7.3.1. □

Theorem 8.3.2. *$S_{\mathrm{Randles,s}}(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n; \varphi_1, \varphi_2)$ is affine invariant for $n > p(p-1)$.*

Proof. For a transformation with a nonsingular $p \times p$ matrix D, we use the notation from the proof of Theorem 8.2.1. Further, in the space of transformed observations, we use a superscript D for the angles φ_i and the corresponding vectors $\boldsymbol{u}_i$, $\tilde{\boldsymbol{u}}_{i,n}$ and $\boldsymbol{w}_{i,n}$.

$$\tilde{\boldsymbol{u}}^{\mathrm{D}}_{i,n} = \widehat{\mathrm{A}}_{\mathrm{d},\boldsymbol{Y}}\boldsymbol{u}^{\mathrm{D}}_i = \widehat{\mathrm{A}}_{\mathrm{d},\boldsymbol{Y}}\,\mathrm{sgn}(\mathrm{D}\boldsymbol{u}_i) = a_i\widehat{\mathrm{A}}_{\mathrm{d},\boldsymbol{Y}}\mathrm{D}\boldsymbol{u}_i = a_i c_0^{1/2}\Theta\widehat{\mathrm{A}}_{\mathrm{d},\boldsymbol{X}}\boldsymbol{u}_i = a_i c_0^{1/2}\Theta\tilde{\boldsymbol{u}}_{i,n},$$

where a_i is a positive constant. Therefore, $\boldsymbol{w}^{\mathrm{D}}_{i,n} = \Theta\boldsymbol{w}_{i,n}$. Using $\bar{\boldsymbol{V}}_{\boldsymbol{Y}} = \Theta\bar{\boldsymbol{V}}_{\boldsymbol{X}}$,

$$\begin{aligned}
&S_{\mathrm{Randles,s}}(\mathrm{D}\boldsymbol{X}_1, \ldots, \mathrm{D}\boldsymbol{X}_n; \varphi^{\mathrm{D}}_1, \varphi^{\mathrm{D}}_2)\\
&\quad= \min\left(-\sqrt{2n}\left(\boldsymbol{w}^{\mathrm{D}}_{1,n}\right)^{\mathrm{T}}\bar{\boldsymbol{V}}_{\boldsymbol{Y}}, -\sqrt{2n}\left(\boldsymbol{w}^{\mathrm{D}}_{2,n}\right)^{\mathrm{T}}\bar{\boldsymbol{V}}_{\boldsymbol{Y}}\right)\\
&\quad= \min\left(-\sqrt{2n}\boldsymbol{w}^{\mathrm{T}}_{1,n}\Theta^{\mathrm{T}}\Theta\bar{\boldsymbol{V}}_{\boldsymbol{X}}, -\sqrt{2n}\boldsymbol{w}^{\mathrm{T}}_{2,n}\Theta^{\mathrm{T}}\Theta\bar{\boldsymbol{V}}_{\boldsymbol{X}}\right)\\
&\quad= S_{\mathrm{Randles,s}}(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n; \varphi_1, \varphi_2)
\end{aligned}$$

because of the orthogonality of Θ.

Translation invariance is again given by definition (see Section 2.4). □

Theorem 8.3.3. *Under H_0, $S_{\mathrm{Randles,s}}(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n; \varphi_1, \varphi_2)$ is distribution-free for distributions of the $\boldsymbol{X}_i$'s belonging to the elliptical directions class.*

Proof. We can write the $\boldsymbol{X}_i$'s coming from a distribution with elliptical directions as $\boldsymbol{X}_i = R_i\mathrm{B}^{-1}\boldsymbol{Y}_i$, where $\boldsymbol{Y}_i$ is uniformly distributed on the unit circle, B is a nonsingular matrix and the R_i's are (random or fixed) positive scalars. These scalars have no influence on the test statistic. Further, we have shown in Theorem 8.3.2 that the test statistic is affine invariant, such that its distribution does not depend on B. □

Theorem 8.3.3 allows for the tabulation of critical values for the test based on $S_{\mathrm{Randles,s}}$ when the $\boldsymbol{X}_i$'s are from a directionally symmetric distribution. A permutation test is also possible for directionally symmetric distributions; note that $\widehat{\mathrm{A}}_{\mathrm{d}}$ does not change if some of the observations are multiplied by -1. (Similar observations were made by Randles, 2000, for his unrestricted test based on Q_{d}, and they also apply to $S_{\mathrm{Randles,d}}$).

Examples can easily be found that show that $S_{\mathrm{Randles,s}}(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n; \varphi_1, \varphi_2)$ and the tests based on it are not cone order monotone in the sample with respect to Θ_1 or Θ_1^* (and therefore not cone order monotone in each observation either).

Regarding generalizations to the multivariate case, the same remarks hold as those made for the spatial sign test at the end of Section 7.3.

Chapter 9

"Hodges-Type" Tests

9.1 Hodges' Bivariate Sign Test

One of the earliest proposals of a bivariate sign test for $H_0 : \boldsymbol{\vartheta} = \mathbf{0}$ against the unrestricted shift alternative $H_1 : \boldsymbol{\vartheta} \neq \mathbf{0}$ is by Hodges (1955). He uses the test statistic

$$S_{\text{Hodges}}(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n) = \sup_{\|\boldsymbol{a}\|=1} \#\{i : \boldsymbol{a}^{\mathrm{T}} \boldsymbol{X}_i > 0\},$$

which is the supremum of the univariate sign test statistics on all possible projections of the data. For this test statistic, Hodges derives a formula for the upper tail of the conditional null distribution given the lines through the origin that the observations lie on. Hodges mentions that the statistic is also applicable for higher dimensions than $p = 2$, but that the distribution will not be easily tractable.

Essentially the same test is proposed in the context of circular distributions by Ajne (1968), who derives the formula for the entire conditional null distribution. The test is therefore also known as the *Hodges–Ajne test*.

We use the term *"Hodges-type" tests* here for tests that are based on the supremum or infimum of univariate sign test statistics over a certain range of projections of the data – not only in the bivariate case, but also in higher dimensions.

9.2 Sign Test by Larocque and Labarre (2004)

Larocque and Labarre (2004) propose a conditionally distribution-free sign test for $H_0 : \boldsymbol{\vartheta} = \mathbf{0}$ vs. $H_1 : \exists\, i : \vartheta_i > 0$. (The authors use two contradictory definitions of the hypotheses, but the test seems to be more appropriate for the one given here.) Larocque and Labarre modify Hodges' test statistic by restricting the

direction vectors $\boldsymbol{a}$ over which the supremum is taken to the ones in the positive orthant, i. e.

$$S_{\text{LarLab}}(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n) = \sup_{\|\boldsymbol{a}\|=1, \boldsymbol{a} \geq \boldsymbol{0}} \#\{i : \boldsymbol{a}^{\mathrm{T}} \boldsymbol{X}_i > 0\}.$$

This test statistic is shown to be conditionally distribution-free under H_0 for some class of distributions including the continuous directionally symmetric distributions, where the conditioning is with respect to the number M of observations outside the positive and the negative orthant; for dimensions $p > 2$, an additional, more technical conditioning is used. For $p = 2$, the conditional distribution can be explicitly obtained, while for $p > 2$, it has to be approximated by simulation.

9.3 Modification for a One-Orthant Alternative

The supremum in the definition of S_{LarLab} is taken over the positive orthant only. However, it will still lead to significant test results if only one component of the observations has sufficiently many positive values, even though all other components may always have distinctly negative values. This is not desirable for a one-orthant alternative with a composite null hypothesis, where we would like to conclude from a significant test result that the symmetry center of the distribution is positive in all components.

If, in contrast, we replace the supremum by an infimum, i. e.

$$S_{\text{LarLab,o}}(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n) = \inf_{\|\boldsymbol{a}\|=1, \boldsymbol{a} \geq \boldsymbol{0}} \#\{i : \boldsymbol{a}^{\mathrm{T}} \boldsymbol{X}_i > 0\},$$

the statistic gives the minimal number of points being on the positive side for each projection along a direction in the positive orthant. This is once more in the spirit of an intersection-union test. Since we only use a simple null hypothesis for determining the decision rule, the test will still not respect the level for the (composite) $(2^p - 1)$-orthant null hypothesis, but it will nevertheless be more suitable for this kind of problem than the original test based on S_{LarLab}.

For a directionally symmetric distribution of the observations, the distribution of $S_{\text{LarLab,o}}$ under the simple null hypothesis $\boldsymbol{\vartheta} = \boldsymbol{0}$ is the same as that of $n - S_{\text{LarLab}}$, such that we can derive conditional distributions from Theorem 1 in Larocque and Labarre (2004), or from their Theorem 2 for the special case $p = 2$. (Note that the proof of Theorem 2 given by Larocque and Labarre contains two small mistakes: R_j should be defined as $R_0 + \sum_{k=1}^{j} R_k$, and the expression given for $\mathrm{P}\,(Z \geq x \,|M = m\,)$ is valid for every $x \in \mathbb{Z}$ – it is not only used for the values indicated.)

9.4 Extension to a Convex Cone Alternative

We can adapt the test statistic to the more general alternative $H_1 : \boldsymbol{\vartheta} \in C$, where $C \subset \mathbb{R}^p$ is a convex cone. For this situation, we can examine whether a sufficiently high proportion of the data is on the "right" side of each hyperplane tangent to the cone (i. e. on the same side as the cone):

$$S_{\text{LarLab,c}}(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n; C) = \inf_{\|\boldsymbol{a}\|=1, \boldsymbol{a} \in C^*} \#\{i : \boldsymbol{a}^{\mathrm{T}} \boldsymbol{X}_i > 0\},$$

where $C^* = \{\boldsymbol{a} : \boldsymbol{a}^{\mathrm{T}} \boldsymbol{c} \geq 0 \; \forall \, \boldsymbol{c} \in C\}$ is the positive dual of the cone C. (If C is the positive orthant, the statistic reduces to $S_{\text{LarLab,o}}(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)$ because $C^* = C$.)

We can derive the conditional null distribution of this statistic in analogy to that of $S_{\text{LarLab,o}}$. In order to adapt the results in Larocque and Labarre (2004), we modify the notation as follows:

$$\begin{aligned} M &= \#\{i : \boldsymbol{X}_i \notin C \cup -C\} \\ Y &= \#\{i : \boldsymbol{X}_i \in C\} \end{aligned}$$

We only investigate in detail the conditional distribution of $S_{\text{LarLab,c}}$ given M in the bivariate case here. In agreement with the notation used before, we denote the statistic $S_{\text{LarLab,c}}$ as $S_{\text{LarLab,s}}$ in this case of a sector alternative. The adaptation of Theorem 2 in Larocque and Labarre (2004) can be carried out as follows:

Theorem 9.4.1. *Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be a sample from a bivariate distribution that is directionally symmetric with respect to $\mathbf{0}$. Let C be the sector between the angles φ_1 and φ_2, with $\varphi_2 - \varphi_1 \in (0, \pi)$. For M as above and $s \in \{0, \ldots, n - \lceil m/2 \rceil\}$,*

$$\mathrm{P}\left(S_{\text{LarLab,s}}(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n; \varphi_1, \varphi_2) \leq s \,|M = m\right) = \frac{1}{2^n} \sum_{y=0}^{n-m} F(m - s + y) \binom{n-m}{y},$$

where

$$\begin{aligned} F(x) &= \sum_{\substack{a=0 \\ x < \max(a, m-a)}}^{m} \binom{m}{a} + \sum_{\substack{a=0 \\ \max(a, m-a) \leq x \leq m}}^{m} \binom{m}{x} \\ &= 1(x \leq m) \sum_{a=0}^{m} \binom{m}{\max(x, a, m-a)} \end{aligned}$$

for $x \in \mathbb{Z}$.

Proof. Let $\xi_i \in [0, 2\pi)$ be the (oriented) angle between the second coordinate axis and $\boldsymbol{X}_i$, and $\tilde{\xi}_i = \xi_i \bmod \pi \in [0, \pi)$, which is equal to the (oriented) angle

between the first coordinate axis and the line perpendicular on $\boldsymbol{X}_i$. Without loss of generality, assume $\varphi_2 = \frac{\pi}{2}$ and $0 < \tilde{\xi}_1 < \ldots < \tilde{\xi}_n < \pi$.

There are

$$M = \#\{i : \boldsymbol{X}_i \notin C \cup -C\} = \#\left\{i : 0 \leq \tilde{\xi}_i < \varphi_1 + \frac{\pi}{2}\right\}$$

observations that do not have the same sign for every projection examined. For these M observations (i. e. for $i = 1, \ldots, M$), let

$$Z_i = \begin{cases} -1 & \text{if } \xi_i \geq \pi, \\ +1 & \text{if } \xi_i < \pi. \end{cases}$$

(Z_i is the change in the number of observations with a positive projection when the direction of the projection passes through ξ_i counterclockwise.) These random variables are independent with $\mathrm{P}(Z_i = -1) = \mathrm{P}(Z_i = +1) = 0.5$. Therefore,

$$R_j^* = \sum_{i=1}^{j} Z_i \quad (j = 0, \ldots, M)$$

is a symmetric binary random walk on $\mathbb{Z}$ starting at $R_0^* = 0$, for which by symmetry and Theorem B.4.1 (for a fixed $M = m$)

$$\begin{aligned} \mathrm{P}\left(\min_{0 \leq j \leq m} R_j^* \leq b \,\middle|\, R_m^* = m - 2a\right) &= \mathrm{P}\left(\max_{0 \leq j \leq m} R_j^* \geq -b \,\middle|\, R_m^* = 2a - m\right) \\ &= \frac{\binom{m}{(m-2b-2a+m)/2}}{\binom{m}{(m+2a-m)/2}} = \frac{\binom{m}{m-a-b}}{\binom{m}{a}} \end{aligned}$$

for $a \in \{0, \ldots, m\}$ and $b \in \{-a, \ldots, \min(0, m - 2a)\}$.

Now, let

$$\begin{aligned} R_0 &= \#\{i : \pi \leq \xi_i < \frac{3\pi}{2} + \varphi_1\} = \#\{i \leq M : Z_i = -1\}, \\ R_j &= R_0 + R_j^* \quad (j = 1, \ldots, M) \end{aligned}$$

be a shifted version of this random walk. By the definition of R_0, $R_M = R_0 + (M - R_0) - R_0 = M - R_0$. For $r_0 \in \{0, \ldots, m\}$,

$$\begin{aligned} &\mathrm{P}\left(\min_{0 \leq j \leq m} R_j \leq b \,\middle|\, M = m, R_0 = r_0\right) \\ &= \mathrm{P}\left(\min_{0 \leq j \leq m} R_j^* \leq b - r_0 \,\middle|\, M = m, R_m^* = m - 2r_0\right) \\ &= \begin{cases} 0 & \text{for } b < 0, \\ \frac{\binom{m}{m-r_0-(b-r_0)}}{\binom{m}{r_0}} = \frac{\binom{m}{m-b}}{\binom{m}{r_0}} & \text{for } b \in \{0, \ldots, \min(r_0, m - r_0)\}, \\ 1 & \text{for } b > \min(r_0, m - r_0). \end{cases} \end{aligned}$$

Given $M = m$, R_0 has a binomial distribution with parameters m and 0.5, and therefore

$$\begin{aligned}
&\mathrm{P}\left(\min_{0\le j\le m} R_j \le b \,\middle|\, M = m\right) \\
&\quad= \sum_{r_0=0}^{m} \mathrm{P}\left(R_0 = r_0 \,|M = m\right) \mathrm{P}\left(\min_{0\le j\le m} R_j \le b \,\middle|\, M = m, R_0 = r_0\right) \\
&\quad= \sum_{\substack{r_0=0 \\ \min(r_0, m-r_0)\ge b}}^{m} \frac{1}{2^m}\binom{m}{r_0}\frac{\binom{m}{m-b}}{\binom{m}{r_0}} + \sum_{\substack{r_0=0 \\ \min(r_0, m-r_0)< b}}^{m} \frac{1}{2^m}\binom{m}{r_0} \\
&\quad= \frac{1}{2^m}\left[\sum_{\substack{r_0=0 \\ \max(r_0, m-r_0)\le m-b\le m}}^{m} \binom{m}{m-b} + \sum_{\substack{r_0=0 \\ m-b<\max(r_0, m-r_0)}}^{m} \binom{m}{r_0}\right] \\
&\quad= \frac{1}{2^m} F(m-b),
\end{aligned}$$

for $b \ge 0$ (this conditional probability is 0 for $b < 0$).

We define the angles θ_j such that $0 < \theta_0 < \tilde{\xi}_1 < \theta_1 < \ldots < \theta_{M-1} < \tilde{\xi}_M < \theta_M < \varphi_1 + \frac{\pi}{2}$ and the corresponding unit vectors $\boldsymbol{a}_j = (\cos\theta_j, \sin\theta_j)^{\mathrm{T}}$. Looking at the random walk given by $\#\{i : \boldsymbol{a}_j^{\mathrm{T}}\boldsymbol{X}_i > 0\}$ $(j = 0, \ldots, M)$, we can write

$$\begin{aligned}
S_{\mathrm{LarLab,s}}(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n; \varphi_1, \varphi_2) &= \inf_{\|\boldsymbol{a}\|=1, \boldsymbol{a}\in C^*} \#\{i : \boldsymbol{a}^{\mathrm{T}}\boldsymbol{X}_i > 0\} \\
&= \min_{0\le j\le M} \#\left\{i : \boldsymbol{a}_j^{\mathrm{T}}\boldsymbol{X}_i > 0\right\} \\
&= \min_{0\le j\le M} R_j^* + \#\left\{i : X_{i1} > 0\right\} \\
&= \min_{0\le j\le M} R_j^* + R_0 + \#\left\{i : \boldsymbol{X}_i \in C\right\} \\
&= \min_{0\le j\le M} R_j + \#\{i : \boldsymbol{X}_i \in C\}.
\end{aligned}$$

The second term in the last line has a binomial distribution with parameters $n-m$ and 0.5 (given $M = m$). The two terms are independent (given $M = m$), and we can therefore obtain the conditional distribution of the sum by the following

convolution:

$$\begin{aligned}
&\mathrm{P}\left(S_{\mathrm{LarLab,s}}(\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n;\varphi_1,\varphi_2)\le s\,|M=m\right)\\
&=\sum_{y=0}^{n-m}\mathrm{P}\left(\min_{0\le j\le m}R_j\le s-y\,\middle|M=m\right)\mathrm{P}\left(\#\{i:\boldsymbol{X}_i\in C\}=y\,|M=m\right)\\
&=\sum_{y=0}^{n-m}\frac{1}{2^m}\,F(m-s+y)\frac{1}{2^{n-m}}\binom{n-m}{y}\\
&=\frac{1}{2^n}\sum_{y=0}^{n-m}F(m-s+y)\binom{n-m}{y}
\end{aligned}$$

□

For the discussion of further properties, we return to the general p-variate case and to the test for a convex cone alternative.

Theorem 9.4.2. *$S_{\mathrm{LarLab,c}}(\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n;C)$ is affine invariant.*

Proof. Let D be any nonsingular $p\times p$ matrix. Then

$$\begin{aligned}
S_{\mathrm{LarLab,c}}(\mathrm{D}\boldsymbol{X}_1,\ldots\mathrm{D}\boldsymbol{X}_n;\mathrm{D}C) &= \inf_{\|\boldsymbol{a}\|=1,\boldsymbol{a}\in(\mathrm{D}C)^*}\#\{i:\boldsymbol{a}^{\mathrm{T}}\mathrm{D}\boldsymbol{X}_i>0\}\\
&= \inf_{\|\boldsymbol{a}\|=1,\boldsymbol{a}^{\mathrm{T}}\boldsymbol{c}\ge 0\ \forall\ \boldsymbol{c}\in\mathrm{D}C}\#\{i:\boldsymbol{a}^{\mathrm{T}}\mathrm{D}\boldsymbol{X}_i>0\}\\
&= \inf_{\|\boldsymbol{a}\|=1,\boldsymbol{a}^{\mathrm{T}}\mathrm{D}\boldsymbol{c}\ge 0\ \forall\ \boldsymbol{c}\in C}\#\{i:\boldsymbol{a}^{\mathrm{T}}\mathrm{D}\boldsymbol{X}_i>0\}\\
&= \inf_{\|\tilde{\boldsymbol{a}}\|=1,\tilde{\boldsymbol{a}}^{\mathrm{T}}\boldsymbol{c}\ge 0\ \forall\ \boldsymbol{c}\in C}\#\{i:\tilde{\boldsymbol{a}}^{\mathrm{T}}\boldsymbol{X}_i>0\}\\
&= \inf_{\|\tilde{\boldsymbol{a}}\|=1,\tilde{\boldsymbol{a}}^{\mathrm{T}}\in C^*}\#\{i:\tilde{\boldsymbol{a}}^{\mathrm{T}}\boldsymbol{X}_i>0\}\\
&= S_{\mathrm{LarLab,c}}(\boldsymbol{X}_1,\ldots\boldsymbol{X}_n;C).
\end{aligned}$$

For the fourth equality, we have defined $\tilde{\boldsymbol{a}}=\mathrm{D}^{\mathrm{T}}\boldsymbol{a}$. (Note that the condition $\|\boldsymbol{a}\|=1$ for the infimum is equivalent to $\|\boldsymbol{a}\|>0$ – it is just used for selecting one representative for each direction. It can therefore also be replaced by $\|\tilde{\boldsymbol{a}}\|>0$ or $\|\tilde{\boldsymbol{a}}\|=1$.)

Translation invariance is given by definition (see Section 2.4). □

Because the quantities used for conditioning (M, Y, but also the more complicated properties used for $p>2$) are invariant under affine transformations, the result of the test based on the conditional distribution of $S_{\mathrm{LarLab,c}}$ under $\boldsymbol{\vartheta}=\mathbf{0}$ is also affine invariant.

The test statistic $S_{\mathrm{LarLab,c}}$ is cone order monotone in each observation with respect to C: For $\boldsymbol{a}\in C^*$ and $\boldsymbol{c}\in C$, $\boldsymbol{a}^{\mathrm{T}}(\boldsymbol{X}_i+\boldsymbol{c})=\boldsymbol{a}^{\mathrm{T}}\boldsymbol{X}_i+\boldsymbol{a}^{\mathrm{T}}\boldsymbol{c}\ge\boldsymbol{a}^{\mathrm{T}}\boldsymbol{X}_i$ by the definition of the positive dual.

However, there are examples showing that the tests based on the conditional null distribution of $S_{\text{LarLab,c}}$ are not cone order monotone in the sample with respect to C or C^* (and therefore not cone order monotone in each observation either). Such examples occur when a translation of the observations does not change $S_{\text{LarLab,c}}$ but does reduce M.

Chapter 10

Graphical Comparison Methods

Given the availability of several test proposals, we need methods that allow for the selection of an appropriate test in a specific situation, i. e. of a test that is suitable for a given combination of null and alternative hypotheses. Further, we would like to compare different tests.

In this chapter, we present three methods that allow for a graphical assessment of the properties of a test. The first two sections deal with such methods from the literature. In Section 10.3, we propose a novel approach.

We present these methods using simple, highly parametric examples, for which most of the required distributions are readily available. In Chapter 11, we will apply two of these comparison methods to some of the nonparametric tests discussed in the previous chapters.

10.1 Acceptance and Rejection Regions

A method that is commonly used to visualize bivariate location tests in the parametric context is that of drawing the border of the acceptance region, i. e. of the set of values of some (bivariate) statistic that lead to the acceptance of the null hypothesis. (The complement of this set is the rejection region.) If the statistic used is the same for different test procedures, this method also allows for a comparison of the procedures. Such plots can be found, e. g., in Follmann (1996) and Logan (2003).

For illustration purposes, we use a very simple situation:

Definition 10.1.1. Assume that some statistic $\boldsymbol{Z}$ has a bivariate normal distribution with the covariance matrix known to be the identity matrix but with unknown location parameter $\boldsymbol{\vartheta}$. Consider the simple $H_0 : \boldsymbol{\vartheta} = \mathbf{0}$, and choose $\alpha \in (0, \frac{1}{2})$. Define tests using the following decision rules based on a single observed value $\boldsymbol{Z} = \boldsymbol{z}$:

(a) *Unrestricted χ^2 test*: Reject H_0 for $z_1^2 + z_2^2 > \chi^2_{1-\alpha}(2)$.

(b) *z min test for the simple* H_0: Reject H_0 for $\min(z_1, z_2) > z_{1-\sqrt{\alpha}}$.

(c) *z max test*: Reject H_0 for $\max(z_1, z_2) > z_{\sqrt{1-\alpha}}$.

(d) *z sum test*: Reject H_0 for $z_1 + z_2 > \sqrt{2}\, z_{1-\alpha}$.

(e) X_+^2 *test by Follmann (1996)*: Reject H_0 for $z_1^2 + z_2^2 > \chi^2_{1-2\alpha}(2)$ and, simultaneously, $z_1 + z_2 > 0$.

(For $0 < p < 1$, z_p and $\chi^2_p(k)$ denote the p quantile of the standard normal distribution and of the chi-square distribution with k degrees of freedom, respectively.) △

Note that due to the assumption of independent components and the simple null hypothesis, the z max test and the z min test can be formulated in a more powerful way than would result from the considerations in Sections 5.1 and 5.2.

Theorem 10.1.1. *The tests from Definition 10.1.1 (a)–(e) are of level* α *for the simple* $H_0 : \boldsymbol{\vartheta} = \mathbf{0}$.

Proof.

(a) $Z_1^2 + Z_2^2$ has a χ^2 distribution with 2 degrees of freedom.

(b)
$$\begin{aligned} \mathrm{P}_{\mathbf{0}}\left(\min(Z_1, Z_2) > z_{1-\sqrt{\alpha}}\right) &= \mathrm{P}_{\mathbf{0}}\left(Z_1 > z_{1-\sqrt{\alpha}}\right) \mathrm{P}_{\mathbf{0}}\left(Z_2 > z_{1-\sqrt{\alpha}}\right) \\ &= \left(\sqrt{\alpha}\right)^2 = \alpha. \end{aligned}$$

(c)
$$\begin{aligned} \mathrm{P}_{\mathbf{0}}\left(\max(Z_1, Z_2) > z_{\sqrt{1-\alpha}}\right) &= 1 - \mathrm{P}_{\mathbf{0}}\left(Z_1 \le z_{\sqrt{1-\alpha}}\right) \mathrm{P}_{\mathbf{0}}\left(Z_2 \le z_{\sqrt{1-\alpha}}\right) \\ &= 1 - \left(\sqrt{1-\alpha}\right)^2 = \alpha. \end{aligned}$$

(d) $Z_1 + Z_2$ has an $\mathrm{N}(0, 2)$ distribution, of which $\sqrt{2}\, z_{1-\alpha}$ is the $1-\alpha$ quantile.

(e) Due to the symmetry of the distribution of $\boldsymbol{Z}$ with respect to the line $Z_1 + Z_2 = 0$,

$$\begin{aligned} \mathrm{P}_{\mathbf{0}}\left(Z_1^2 + Z_2^2 > \chi^2_{1-2\alpha}(2), Z_1 + Z_2 > 0\right) &= \frac{1}{2}\, \mathrm{P}_{\mathbf{0}}\left(Z_1^2 + Z_2^2 > \chi^2_{1-2\alpha}(2)\right) \\ &= \frac{1}{2} \cdot 2\alpha = \alpha. \end{aligned}$$

□

The rejection regions of these test procedures for $\alpha = 0.05$ are shown in Figure 10.1; they were determined as follows:

(a) Unrestricted χ^2 test: $\sqrt{z_1^2 + z_2^2} > \sqrt{\chi^2_{0.95}(2)} \approx \sqrt{5.99} \approx 2.45$.

(b) z min test for the simple H_0: $\min(z_1, z_2) > z_{1-\sqrt{0.05}} \approx 0.76$.

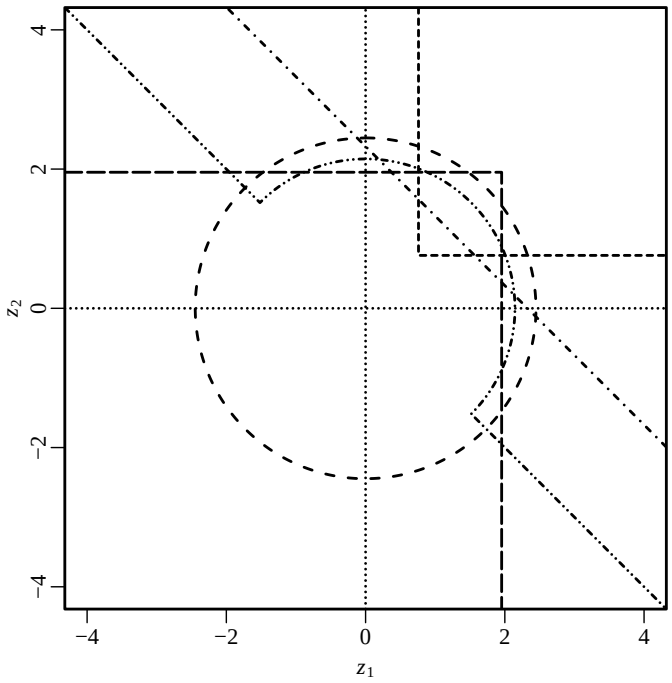

Figure 10.1: Acceptance and rejection regions for a bivariate normal statistic with independent components of variance 1. (The upper right corner of the plot area is in the rejection region for each test.) - - unrestricted χ^2 test; ---- z min test for the simple H_0; —— z max test; -·-· z sum test; ····· Follmann's X_+^2 test.

(c) z max test: $\max(z_1, z_2) > z_{\sqrt{0.95}} \approx 1.95$.

(d) z sum test: $z_1 + z_2 > \sqrt{2}\, z_{0.95} \approx 2.33$.

(e) X_+^2 test by Follmann (1996): $\sqrt{z_1^2 + z_2^2} > \sqrt{\chi^2_{0.9}(2)} \approx \sqrt{4.61} \approx 2.15$ and $z_1 + z_2 > 0$.

From Figure 10.1, we see that the rejection region of each of these five tests is closer to the origin than the other four rejection regions for certain directions. The figure suggests that

- the z min test for the simple H_0 is most appropriate for the alternative that $\boldsymbol{\vartheta}$ is in the first quadrant,
- the z max test is suitable for the alternative that $\boldsymbol{\vartheta}$ is outside the third quadrant, and
- the other two restricted alternative tests are appropriate for a half-plane alternative.

We can only use the approach using acceptance and rejection regions for the comparison of tests based on some common (multivariate, preferably p-variate if $\boldsymbol{\vartheta} \in \mathbb{R}^p$) statistic. Such a common underlying statistic (except the sample itself) is not generally available when we would like to compare parametric and nonparametric tests or nonparametric tests among each other, and therefore, this approach is not applicable.

A further, inherent problem of the approach is that, in some sense, the assessment of tests is done in the wrong space – the observed properties (in the space of the statistic) have to be reinterpreted in the parameter space. As an example, it is impossible to conclude from Figure 10.1 that the z min test for the simple H_0 is actually liberal for the three-quadrant null hypothesis, which is the case since the level α is only controlled at the origin. (A more general, but rather conservative version of the z min test that is actually of level α for the three-quadrant null hypothesis can be defined by rejecting H_0 for $\min(z_1, z_2) > z_{1-\alpha}$; see Section 5.2.)

10.2 Power at a Fixed Distance from the Origin

We can also examine a test graphically in the bivariate case by plotting the curve of the power depending on the angle of $\boldsymbol{\vartheta}$, while the distance of $\boldsymbol{\vartheta}$ from the origin is held constant. A plot of the power curves of several tests allows for an immediate comparison of the powers for any angle (at the given distance from the origin).

As a first example, we calculate the power curves for the same tests as in the preceding section.

Theorem 10.2.1. *The power of the tests from Definition 10.1.1 at some parameter value $\boldsymbol{\vartheta}$ can be obtained as follows:*

(a) *Unrestricted χ^2 test:*

$$\mathrm{P}_{\boldsymbol{\vartheta}}\left(Z_1^2 + Z_2^2 > \chi^2_{1-\alpha}(2)\right) = \mathrm{P}\left(X^2_{2,\|\boldsymbol{\vartheta}\|^2} > \chi^2_{1-\alpha}(2)\right),$$

where $X^2_{2,\|\boldsymbol{\vartheta}\|^2}$ is a noncentral χ^2 random variable with 2 degrees of freedom and noncentrality parameter $\|\boldsymbol{\vartheta}\|^2$.

(b) *z min test for the simple H_0:*

$$\mathrm{P}_{\boldsymbol{\vartheta}}\left(\min(Z_1, Z_2) > z_{1-\sqrt{\alpha}}\right) = \left[1 - \Phi(z_{1-\sqrt{\alpha}} - \vartheta_1)\right]\left[1 - \Phi(z_{1-\sqrt{\alpha}} - \vartheta_2)\right].$$

(c) *z max test:*

$$\mathrm{P}_{\boldsymbol{\vartheta}}\left(\max(Z_1, Z_2) > z_{\sqrt{1-\alpha}}\right) = 1 - \Phi(z_{\sqrt{1-\alpha}} - \vartheta_1)\Phi(z_{\sqrt{1-\alpha}} - \vartheta_2).$$

(d) *z sum test:*

$$\mathrm{P}_{\boldsymbol{\vartheta}}\left(Z_1 + Z_2 > \sqrt{2}\, z_{1-\alpha}\right) = 1 - \Phi\left(z_{1-\alpha} - \frac{\vartheta_1 + \vartheta_2}{\sqrt{2}}\right).$$

(e) *X^2_+ test by Follmann (1996):*

$$\begin{aligned}
&\mathrm{P}_{\boldsymbol{\vartheta}}\left(Z_1^2 + Z_2^2 > \chi^2_{1-2\alpha}(2), Z_1 + Z_2 > 0\right) \\
&\quad = \Phi\left(\tilde{\vartheta}_1\right) - \lim_{k\to\infty} \sum_{j=1}^{k} \left[\Phi\left(a_{k,j} - \tilde{\vartheta}_1\right) - \Phi\left(a_{k,j-1} - \tilde{\vartheta}_1\right)\right] \\
&\qquad\qquad \cdot \left[\Phi\left(a_{k,k+1-j} - \tilde{\vartheta}_2\right) - \Phi\left(-a_{k,k+1-j} - \tilde{\vartheta}_2\right)\right],
\end{aligned}$$

where

$$\begin{aligned}
\tilde{\vartheta}_1 &= \frac{1}{\sqrt{2}}(\vartheta_1 + \vartheta_2), \\
\tilde{\vartheta}_2 &= \frac{1}{\sqrt{2}}(-\vartheta_1 + \vartheta_2), \\
a_{k,j} &= \sin\left(\frac{j}{k} \cdot \frac{\pi}{2}\right)\sqrt{\chi^2_{1-2\alpha}(2)}.
\end{aligned}$$

(Φ denotes the cumulative distribution function of the standard normal distribution.)

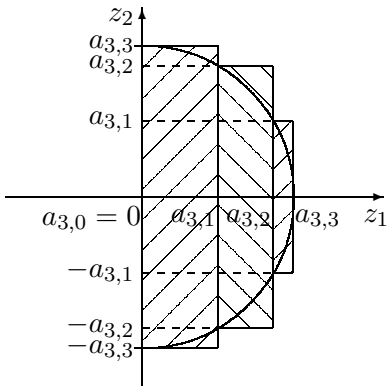

Figure 10.2: Approximation of the region $\{\boldsymbol{z} : z_1^2 + z_2^2 \le \chi^2_{1-2\alpha}(2), z_1 > 0\}$ by $k = 3$ rectangles.

Proof.

(a) $\mathrm{P}_{\boldsymbol{\vartheta}}\left(Z_1^2 + Z_2^2 > \chi^2_{1-\alpha}(2)\right) = \mathrm{P}_{\mathbf{0}}\left((Z_1+\vartheta_1)^2 + (Z_2+\vartheta_2)^2 > \chi^2_{1-\alpha}(2)\right)$, and under $\boldsymbol{\vartheta} = \mathbf{0}$, $(Z_1+\vartheta_1)^2+(Z_2+\vartheta_2)^2$ has a noncentral χ^2 distribution with 2 degrees of freedom and noncentrality parameter $\vartheta_1^2 + \vartheta_2^2 = \|\boldsymbol{\vartheta}\|^2$.

(b)–(d) The proof essentially uses the same decompositions as the one of Theorem 10.1.1.

(e) Using a rotation about the origin by $-\frac{\pi}{4}$, we simplify the desired power to

$$\begin{aligned}&\mathrm{P}_{\tilde{\boldsymbol{\vartheta}}}\left(Z_1^2 + Z_2^2 > \chi^2_{1-2\alpha}(2), Z_1 > 0\right)\\ &\quad = \mathrm{P}_{\tilde{\boldsymbol{\vartheta}}}\left(Z_1 > 0\right) - \mathrm{P}_{\tilde{\boldsymbol{\vartheta}}}\left(Z_1^2 + Z_2^2 \le \chi^2_{1-2\alpha}(2), Z_1 > 0\right), \qquad (*)\end{aligned}$$

where

$$\tilde{\boldsymbol{\vartheta}} = \begin{pmatrix} \cos\left(-\frac{\pi}{4}\right) & -\sin\left(-\frac{\pi}{4}\right) \\ \sin\left(-\frac{\pi}{4}\right) & \cos\left(-\frac{\pi}{4}\right) \end{pmatrix} \boldsymbol{\vartheta} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \boldsymbol{\vartheta}.$$

The first term on the right hand side of $(*)$ is $1 - \Phi(-\tilde{\vartheta}_1) = \Phi(\tilde{\vartheta}_1)$.

Thanks to the rotation performed, we can easily approximate the second term on the right hand side of $(*)$: The probability that $\boldsymbol{Z}$ lies within the half circle $H = \{\boldsymbol{z} : z_1^2 + z_2^2 \le \chi^2_{1-2\alpha}(2), z_1 > 0\}$ is the limit for $k \to \infty$ of the probability that $\boldsymbol{Z}$ lies within one of the rectangles $R_{k,j} = (a_{k,j-1}, a_{k,j}) \times (-a_{k,k+1-j}, a_{k,k+1-j})$ $(j = 1, \ldots, k)$. (In Figure 10.2, the situation is illustrated for $k = 3$.) We can easily express these probabilities using the univariate standard normal distribution function because the components are independent. The area of $\Delta_k = \bigcup_{j=1}^k R_{k,j} \setminus H$ converges to 0, and it follows from the boundedness of the underlying density that the probability of $\boldsymbol{Z}$ lying within Δ_k converges to 0. □

The power curves of the five tests at distance $\sqrt{10}$ from the origin are shown

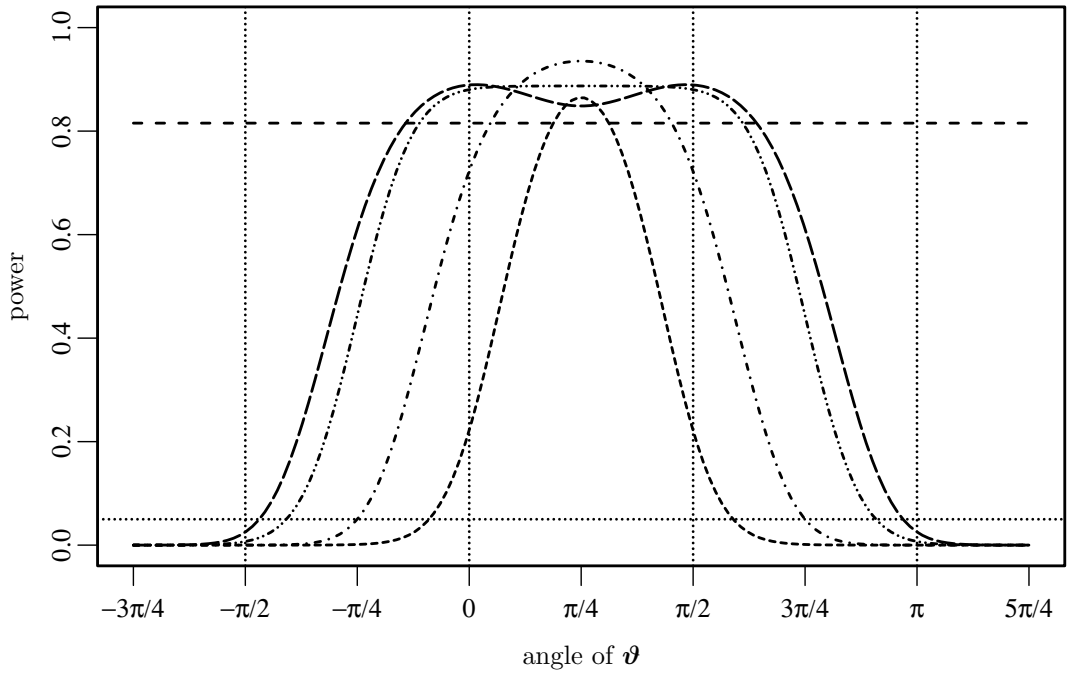

Figure 10.3: Power at distance $\|\boldsymbol{\vartheta}\| = \sqrt{10}$ from the origin for a bivariate normal statistic with independent components of variance 1. The dotted lines indicate the nominal significance level and the quadrant borders, respectively. - - unrestricted χ^2 test; ---- z min test for the simple H_0; —— z max test; -·-· z sum test; -··-·· Follmann's X_+^2 test.

in Figure 10.3. It is obvious that we could restrict our attention to the left or right half of the curves because all tests behave symmetrically with respect to the diagonal at angle $\pi/4$.

From this plot, it is now obvious that the z min test for the simple H_0 is liberal if it is applied to the three-quadrant null hypothesis. Further, we can see that the z sum test has power α for the angles $-\pi/4$ and $3\pi/4$. The plot also reveals that Follmann's X_+^2 test is liberal for the half-plane null hypothesis; its curve and the one for the z max test are quite similar.

Except for such simple cases as the ones used here, the power at some selected angles has to be estimated by simulation in practice. An example of such a plot can be found in Follmann (1996).

While the power function approach described in this section acts directly in the parameter space, which facilitates interpretation, and circumvents the necessity of some common underlying statistic, the following problems arise:

- The resulting curve may depend quite heavily on the chosen distance from the origin. When we compare tests, this may lead to contradictory results.

(In the example, for a distance of 1, the curve of Follmann's X_+^2 test is nearly identical to that of the z max test, except for angles around $\pi/4$. For large distances, however, its curve is much more similar to that of the z sum test.)

- We do not obtain any information about the behavior of a test at the origin. Thus, we cannot use the plot for the assessment of the actual significance level at the simple null hypothesis $\boldsymbol{\vartheta} = \mathbf{0}$.

10.3 Curve of Constant Power

The power function of a test for a p-variate location parameter is a function from $\mathbb{R}^p$ to $[0, 1]$. In the bivariate location problem, its graph is therefore a surface in $\mathbb{R}^3$. In Figure 10.4, this graph is shown for the z min test for the simple H_0, again assuming a bivariate normal distribution with unit covariance matrix.

The method described above, which reduces this graph to two dimensions, is the choice of a fixed distance of $\boldsymbol{\vartheta}$ from the origin. This corresponds to an intersection of the power surface with a vertical cylinder centered at the origin.

A second possibility for the reduction to two dimensions is the consideration of a horizontal section through the graph. The resulting curve separates the parameter values where the test has power below the chosen value from those where it has higher power. In more mathematical terms, the curve is the preimage of the chosen power value under the power function of the test considered. (A third possibility is used by Minhajuddin, Frawley, Schucany, and Woodward, 2006: They look at the power curve when the location parameter varies along some straight line through the origin. Several such plots have to be used if the primary interest is in the behavior of a test for different directions.)

The choice of the power value where the horizontal section is taken is arbitrary, but a natural value is given by the desired nominal significance level α of the test. For the situation shown in Figure 10.4, the resulting curve is given in Figure 10.5, together with the curves for the other tests used in this chapter. (The calculation of the curves is described in Theorem 10.3.1 below.)

In the composite null hypothesis case, if the chosen power value is equal to α, the curve should ideally (for an unbiased test) separate Θ_1 (where the power should be at least α) from its complement (the parameter values corresponding to the null hypothesis, for which the power should be at most α). Thus, the curve indicates the shape of the hypothesis combinations that a test may be appropriate for. The distance of the curve from the origin also gives an impression of how well the true significance level matches the nominal one for the simple null hypothesis. We can easily combine several such curves for different tests into one plot for comparison purposes.

However, we will rarely encounter the ideal case of a curve coinciding with the

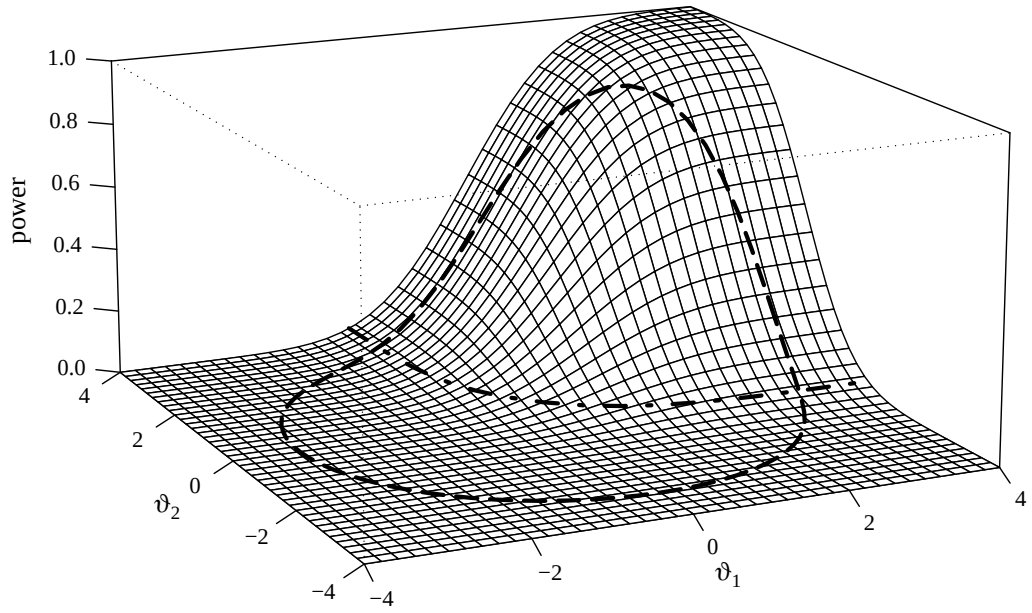

Figure 10.4: The power function of the z min test for a simple H_0 under a bivariate normal distribution with expectation $\boldsymbol{\vartheta}$ and the identity covariance matrix. – – – intersection with a cylinder, i. e. power at some fixed distance (here: $\sqrt{10}$) from the origin; – · – horizontal section, i. e. curve of constant power (here: 0.05).

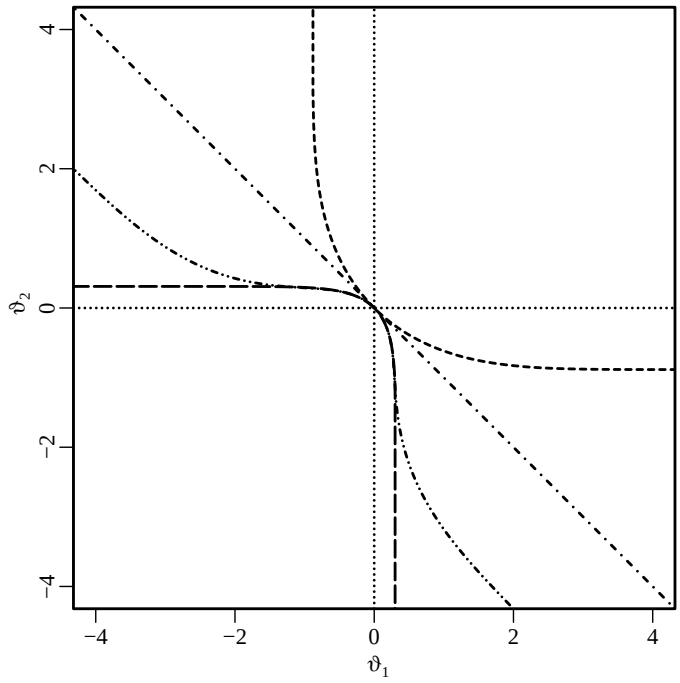

Figure 10.5: Curves of power 0.05 for the four restricted alternative tests for a bivariate normal statistic with independent components of variance 1. The power is above 0.05 to the right of and above each curve. The degenerate curve for the unrestricted χ^2 test would be a point at the origin. ---- z min test for the simple H_0; —— z max test; -·-· z sum test; ····· Follmann's X_+^2 test.

border of Θ_1 in practice: As we have seen in Section 3.2, under some regularity assumptions, there are no unbiased tests for $H_0 : \boldsymbol{\vartheta} \in \mathbb{R}^p \setminus \Theta_1$ vs. $H_1 : \boldsymbol{\vartheta} \in \Theta_1$ when Θ_1 has no smooth border.

The curves of power 0.05 in Figure 10.5 were obtained using the following theorem, except for Follmann's X_+^2 test, where the power was approximated using Theorem 10.2.1 (e) on a grid.

Theorem 10.3.1. *For the tests from Definition 10.1.1, the curves of power β and their asymptotes are given as follows:*

(a) *Unrestricted χ^2 test: The curve with power $\beta \geq \alpha$ is a circle with radius r around the origin; r has to be chosen such that*

$$\mathrm{P}\left(X_{2,r^2}^2 \leq \chi_{1-\alpha}^2(2)\right) = 1 - \beta,$$

where X_{2,r^2}^2 is a noncentral χ^2 random variable with 2 degrees of freedom and noncentrality parameter r^2.

(b) *z min test for the simple H_0:*

$$\vartheta_2 = \vartheta_2(\vartheta_1) = z_{1-\sqrt{\alpha}} - z_{1-\beta}/\left[1-\Phi\left(z_{1-\sqrt{\alpha}}-\vartheta_1\right)\right] \quad (\vartheta_1 > z_{1-\sqrt{\alpha}} - z_{1-\beta})$$
$$\vartheta_2 \to z_{1-\sqrt{\alpha}} - z_{1-\beta} \quad (\vartheta_1 \to \infty)$$

(c) *z max test:*

$$\vartheta_2 = \vartheta_2(\vartheta_1) = z_{\sqrt{1-\alpha}} - z_{(1-\beta)/\Phi\left(z_{\sqrt{1-\alpha}}-\vartheta_1\right)} \quad (\vartheta_1 < z_{\sqrt{1-\alpha}} - z_{1-\beta})$$
$$\vartheta_2 \to z_{\sqrt{1-\alpha}} - z_{1-\beta} \quad (\vartheta_1 \to -\infty)$$

(d) *z sum test:*

$$\vartheta_2 = \vartheta_2(\vartheta_1) = \sqrt{2}\,(z_{1-\alpha} - z_{1-\beta}) - \vartheta_1$$

(e) *X_+^2 test by Follmann (1996):*

$$\vartheta_2 = \vartheta_2(\vartheta_1) \sim \sqrt{2} z_\beta - \vartheta_1 \quad (|\vartheta_1| \to \infty)$$

In (b)–(e), the roles of ϑ_1 and ϑ_2 can be swapped due to the symmetry of the test statistics and of the distribution.

Proof.

(a) From Theorem 10.2.1 (a), it follows that the power depends only on the distance $r = \|\boldsymbol{\vartheta}\|$ from the origin. The condition on r is obvious.

(b)–(d) We equate the expressions for the power given in Theorem 10.2.1 to β, and we can solve these equations for ϑ_2 as a function of ϑ_1 without any difficulty. The asymptotic behavior in (b) and (c) results from the fact that $\Phi(x) \to 0$ for $x \to -\infty$ and $\Phi(x) \to 1$ for $x \to \infty$, respectively.

(e) $|\vartheta_1| \to \infty$ implies $\|\boldsymbol{\vartheta}\| \to \infty$, and therefore

$$\mathrm{P}_{\boldsymbol{\vartheta}}\left(Z_1^2 + Z_2^2 > \chi^2_{1-2\alpha}(2) \,\middle|\, Z_1 + Z_2 > 0\right) \to 1,$$

such that

$$\begin{aligned}
&\mathrm{P}_{\boldsymbol{\vartheta}}\left(Z_1^2 + Z_2^2 > \chi^2_{1-2\alpha}(2), Z_1 + Z_2 > 0\right) \\
&\quad = \mathrm{P}_{\boldsymbol{\vartheta}}\left(Z_1^2 + Z_2^2 > \chi^2_{1-2\alpha}(2) \,\middle|\, Z_1 + Z_2 > 0\right) \mathrm{P}_{\boldsymbol{\vartheta}}\left(Z_1 + Z_2 > 0\right) \\
&\quad \sim \mathrm{P}_{\boldsymbol{\vartheta}}\left(Z_1 + Z_2 > 0\right) = \mathrm{P}_{\boldsymbol{\vartheta}}\left(\frac{Z_1 - \vartheta_1 + Z_2 - \vartheta_2}{\sqrt{2}} > -\frac{\vartheta_1 + \vartheta_2}{\sqrt{2}}\right) \\
&\quad = 1 - \Phi\left(-\frac{\vartheta_1 + \vartheta_2}{\sqrt{2}}\right) = \Phi\left(\frac{\vartheta_1 + \vartheta_2}{\sqrt{2}}\right) \quad (|\vartheta_1| \to \infty).
\end{aligned}$$

We can now equate this expression to β and solve it for ϑ_2. □

For $\alpha = \beta = 0.05$, the asymptotes are $\vartheta_i = z_{1-\sqrt{0.05}} - z_{1-0.05} \approx z_{0.78} - z_{0.95} \approx 0.76 - 1.64 \approx -0.88$ (z min test for the simple H_0), $\vartheta_i = z_{\sqrt{0.95}} - z_{0.95} \approx 1.95 - 1.64 \approx 0.31$ (z max test), $\vartheta_2 = -\vartheta_1$ (z sum test), and $\vartheta_2 = \sqrt{2}\, z_{0.05} - \vartheta_1 \approx 1.41 \cdot (-1.64) - \vartheta_1 \approx -2.33 - \vartheta_1$ (Follmann's X_+^2 test).

In Figure 10.5, we can see that, near the origin, the curves are very similar for the z max test and the X_+^2 test. For the orthant alternatives, none of the curves is ideal, as was to be expected – e. g., the z min test with power 0.05 at the origin has higher power at all other points on the border of the first quadrant. When we think of composite null hypotheses, this test would therefore be more suitable for a non-inferiority setting as was presented in Section 2.3. The z min test according to Section 5.2 would have a curve that asymptotically reaches the border of the first quadrant for $\|\boldsymbol{\vartheta}\| \to \infty$, but this curve would pass noticeably above and to the right of the origin, meaning that the test is conservative near the origin. We will investigate similar curves in the following chapter.

We can also use the concept of curves of parameter values leading to power α to compare the behavior of one test under different distributions. This gives us an impression of how sensitive a test is e. g. to a deviation from a normal distribution.

A short presentation of the approach using curves of constant power is given in Vock (2006). Apparently, the full range of applications of this type of plot has not been recognized before: The only known similar plot in the literature is a contour plot of the power function of a bivariate test in Jennison and Turnbull (1993), which is not suitable for a direct comparison of tests.

For location tests that are also applicable to dimensions $p > 2$, the properties in the bivariate case give a rough indication of what can be expected in higher dimensions. But the method described may also be useful for direct application in higher dimensions: While the graph of the power function would be in $\mathbb{R}^{p+1}$, the preimage of some fixed power value is a subset of $\mathbb{R}^p$. For $p = 3$, this is a surface and can still be visualized, although comparisons are difficult. In higher dimensions, we would have to look at sections.

Chapter 11

Comparison of the Proposed Tests

We now apply the graphical methods from the previous chapter to some of the tests that have been proposed.

The approach comparing acceptance and rejection regions is not used because the tests to be compared are based on entirely different statistics. The power curve at a fixed distance from the origin is only used in one case (uncorrelated normal components) for illustration purposes; we concentrate on the curves of constant power.

The exact and even the asymptotic distributions of the test statistics under specific alternative hypotheses are intractable for most of the tests investigated. However, we can obtain approximate power values using simulations.

Our choice of tests is determined by the following factors: Emphasis is on nonparametric proposals for restricted alternatives, but some parametric tests and tests for non-restricted alternatives are also included for comparison purposes. The proposals have to be precise enough to be implemented. Further, as the comparisons based on simulations are rather time-consuming already for simple test procedures, procedures with a high computational complexity could not be used. For this reason, bootstrap methods are not included. We use the min tests (see Section 5.2) based on the univariate sign and Wilcoxon signed rank test, the cone/opposite cone test (Section 6.2), the sector versions of the spatial sign test by Möttönen and Oja (Section 7.3) and of Randles's test (Section 8.3), the test by Larocque and Labarre (2004), and its version for a convex sector (Sections 9.2 and 9.4). We also include the angle test by Brown (1983) and a one-sample analogue of the multi-sample test by O'Brien (1984), i. e. a Wilcoxon signed rank test applied to the sum of componentwise ranks, subsequently called *O'Brien/Wilcoxon test*. As examples of parametric restricted alternative tests, we use the test based on the conservative bound for the likelihood ratio test statistic by Perlman (1969), the min test based on univariate t tests, and the proposal by Glimm, Srivastava, and Läuter (2002). For the power curve at a fixed distance from the origin, we

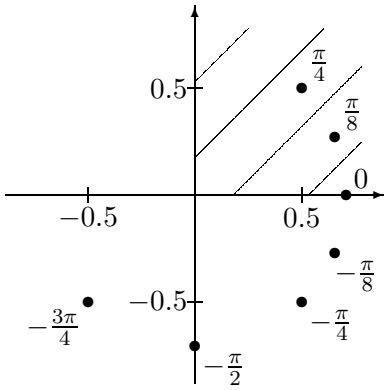

Figure 11.1: Location parameters for the estimation of the power curve at a fixed distance from the origin. The alternative region for most of the tests is the first quadrant, which is hatched. All dots have distance $\sqrt{2}/2$ from the origin, and the respective angle is indicated in the figure.

also compare these tests to tests for unrestricted alternatives, namely the spatial sign test by Möttönen and Oja (1995), the test by Randles (2000), the one by Blumen (1958), and the parametric T^2 test by Hotelling (1931).

In doing power comparisons, of course, the choice of the alternative region, of the underlying distribution, of the sample size, and of the points where the power is calculated is somewhat arbitrary. For the tests that allow for general sector alternative hypotheses, we use the first quadrant as the alternative region; for tests about a specific direction, this hypothetical direction is chosen at the angle $\pi/4$. We use a moderate sample size of $n = 20$.

11.1 Uncorrelated Normal Case

We use the bivariate normal distribution to gain a first impression of the behavior of the tests. The normal distribution is also useful for assessing the loss in efficiency with respect to parametric tests in the case where the latter are optimal. The location parameter $\boldsymbol{\vartheta}$ is the mean; we assume the two components to be uncorrelated, each with variance 1.

11.1.1 Power at a Fixed Distance from the Origin

For $\boldsymbol{\vartheta}$, we use points at the angles $\pi/4$, $\pi/8$, 0, $-\pi/8$, $-\pi/4$, $-\pi/2$, and $-3\pi/4$. The distance from the origin is left constant at $\sqrt{2}/2$. These points and the corresponding angles are shown in Figure 11.1. Due to symmetry, we do not need to consider angles in the other half of the plane.

The results for the uncorrelated normal case are shown in Figure 11.2, in

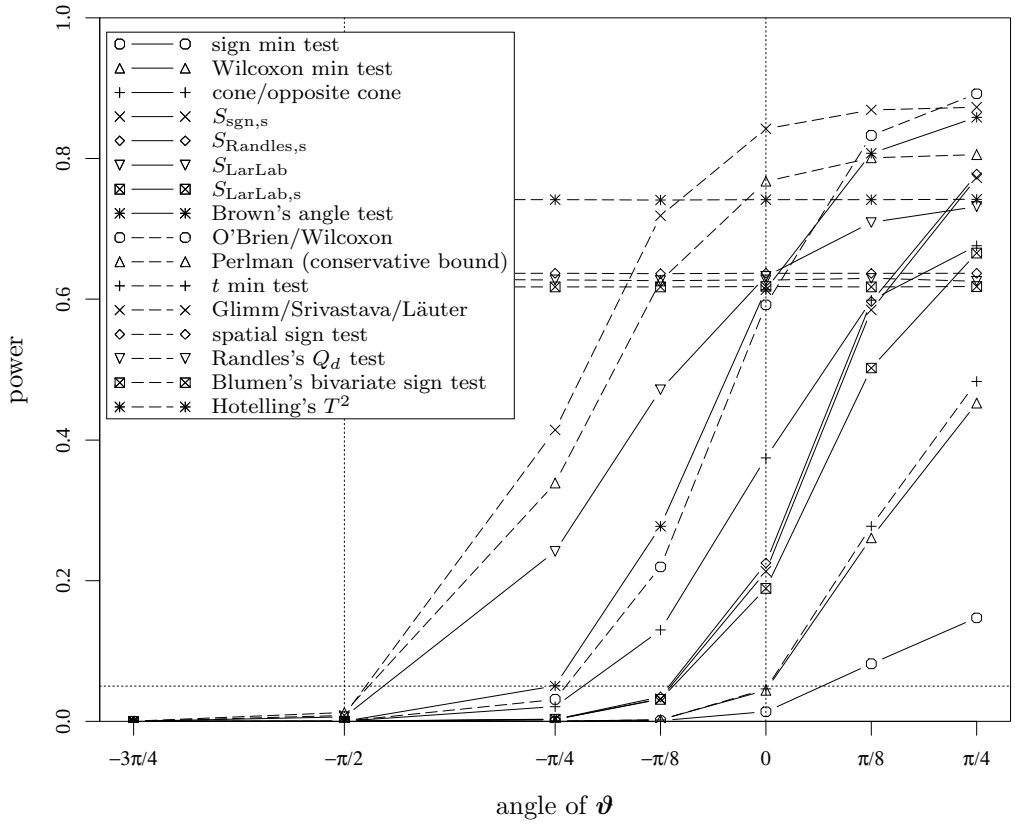

Figure 11.2: Results of the simulation study for the uncorrelated bivariate normal case ($n = 20$): Power of different tests for the alternative values of $\boldsymbol{\vartheta}$ shown in Figure 11.1. The nominal significance level of $\alpha = 0.05$ is indicated by a dotted horizontal line.

which the estimated power (obtained by simulation) is plotted as a function of the angle of $\boldsymbol{\vartheta}$. The points were obtained by at least 100,000 simulations each and are linearly interpolated in the graph.

Based on these results, we can give a first classification of the tests compared:

- *Tests for a point null hypothesis vs. an unrestricted alternative*: The horizontal lines indicate constant power in every direction. This is the case for Blumen's bivariate sign test, the spatial sign test by Möttönen and Oja (1995), the Q_d test by Randles (2000), and the classical T^2 test.

- *Tests for a one-quadrant null hypothesis vs. a three-quadrant alternative*: The power curve of these tests is near zero (and clearly below the significance level) for angles below $-\pi/2$, then rises rapidly. This class contains the test using S_{LarLab} by Larocque and Labarre (2004), the one by Glimm, Srivastava, and Läuter (2002) (originally proposed by the authors for a

one-quadrant alternative), and the procedure using the upper bound for the p-value of the likelihood ratio test by Perlman (1969).

- *Tests for a half-space null hypothesis vs. a half-space alternative*: The power of this class of tests is below the significance level for angles up to $-\pi/4$, then increases and tendentially takes even higher values than for tests in the previous class if alternatives in the first quadrant and close to the main diagonal are investigated. Examples are Brown's angle test, the O'Brien/Wilcoxon test, and, less powerful in the normal case, the cone/opposite cone test.

- *Conservative tests for a three-quadrant null hypothesis vs. a one-quadrant alternative*: For these tests, the power curve exceeds the significance level only for angles larger than 0. These are the min tests based on componentwise sign, Wilcoxon signed rank, or t tests, respectively. (In this case, the t min test corresponds to the likelihood ratio test by Sasabuchi, 1988.) Note that the term "conservative" relates to the behavior for alternatives near the main diagonal, and not for those that are at the border of the positive quadrant but away from the origin.

- *Liberal tests for a three-quadrant null hypothesis vs. a one-quadrant alternative*: In order to avoid the conservativeness of the previous group of tests for alternatives near the main diagonal, these tests are designed to reach the significance level at the origin. As a consequence, they are slightly liberal at the border for the three-quadrant null hypothesis. These tests may be adequate in the case of non-inferiority hypotheses such as those presented in Section 2.3. This class of tests contains the sector variant of the spatial sign test ($S_{\text{sgn,s}}$), the sector variant of Randles's test ($S_{\text{Randles,s}}$), and the one of Larocque and Labarre's test ($S_{\text{LarLab,s}}$).

The classification of the one-sided tests is based on composite null hypotheses. For the simple null hypothesis $H_0 : \boldsymbol{\vartheta} = \mathbf{0}$, all the tests do (at least asymptotically) respect the significance level, such that we can choose a test from the class with highest power for the alternative that we have in mind. According to the figure, this will probably be one of the first three classes mentioned.

11.1.2 Curve of Constant Power

In Figure 11.3 (a), the approximate curves of power 0.05 are shown for the restricted alternative tests from above for the case that the components are uncorrelated and have a standard normal distribution. These curves were estimated by the algorithm described in Appendix C.

The tests for unrestricted alternatives have been omitted in this figure. Their curve with power 0.05 should theoretically be a dot at the origin or, for a conservative test, a small circle around the origin (for a liberal test, there would not exist any such curve at all).

In this figure, we can again identify the classes of tests for restricted alternatives defined above. However, we can make some additional observations:

- The conservative and the liberal tests for a three-quadrant null hypothesis have approximately parallel curves. This can be explained by the fact that these tests are all based on a minimum of two componentwise test statistics (or an infimum over more than two directions in the case of $S_{\text{LarLab,s}}$), but that the critical value is derived from the distribution at the quadrant border (infinitely far away from the origin) or at the origin, respectively. All of these tests are intersection-union tests, but with varying overall level. Actually, in the case of independent components, tests very similar to those of the liberal class with nominal level α can be obtained by performing a test of the conservative class at the nominal level $\sqrt{\alpha}$; vice versa, a test of the liberal class performed at the nominal level α^2 will result in a test similar to those of the conservative class with nominal level α.

- In the class of half-space tests that we defined based on the power at a fixed distance from the origin, the shapes of the curves are quite diverse for the three tests. While the curve of Brown's angle test is approximately a straight line through the origin, the curves of the O'Brien/Wilcoxon test and the cone/opposite cone test are manifestly bent, somewhat resembling the tests for a three-quadrant null hypothesis. These two tests can therefore be regarded as a separate class between the test for a half-space problem (Brown) and the liberal tests for a three-quadrant null hypothesis – for the half-space null and alternative hypotheses, these tests are conservative. (Note that we failed to recognize this difference in the first classification because we only examined the power curve at one fixed distance from the origin. The order of these three tests with respect to the power also depends on the distance from the origin.)

Based on these observations, we can update our classification as follows:

(a) *Tests for a simple null hypothesis vs. an unrestricted alternative*: Blumen's bivariate sign test, spatial sign test by Möttönen and Oja (1995), Q_d test by Randles (2000), T^2 test.

(b) *Tests for a one-quadrant null hypothesis vs. a three-quadrant alternative*: Tests by Larocque and Labarre (2004), Glimm, Srivastava, and Läuter (2002), Perlman (1969).

(c) *Tests for a half-space null hypothesis vs. a half-space alternative*: Brown's angle test.

(d) *Conservative tests for a half-space null hypothesis vs. a half-space alternative*: O'Brien/Wilcoxon test, cone/opposite cone test.

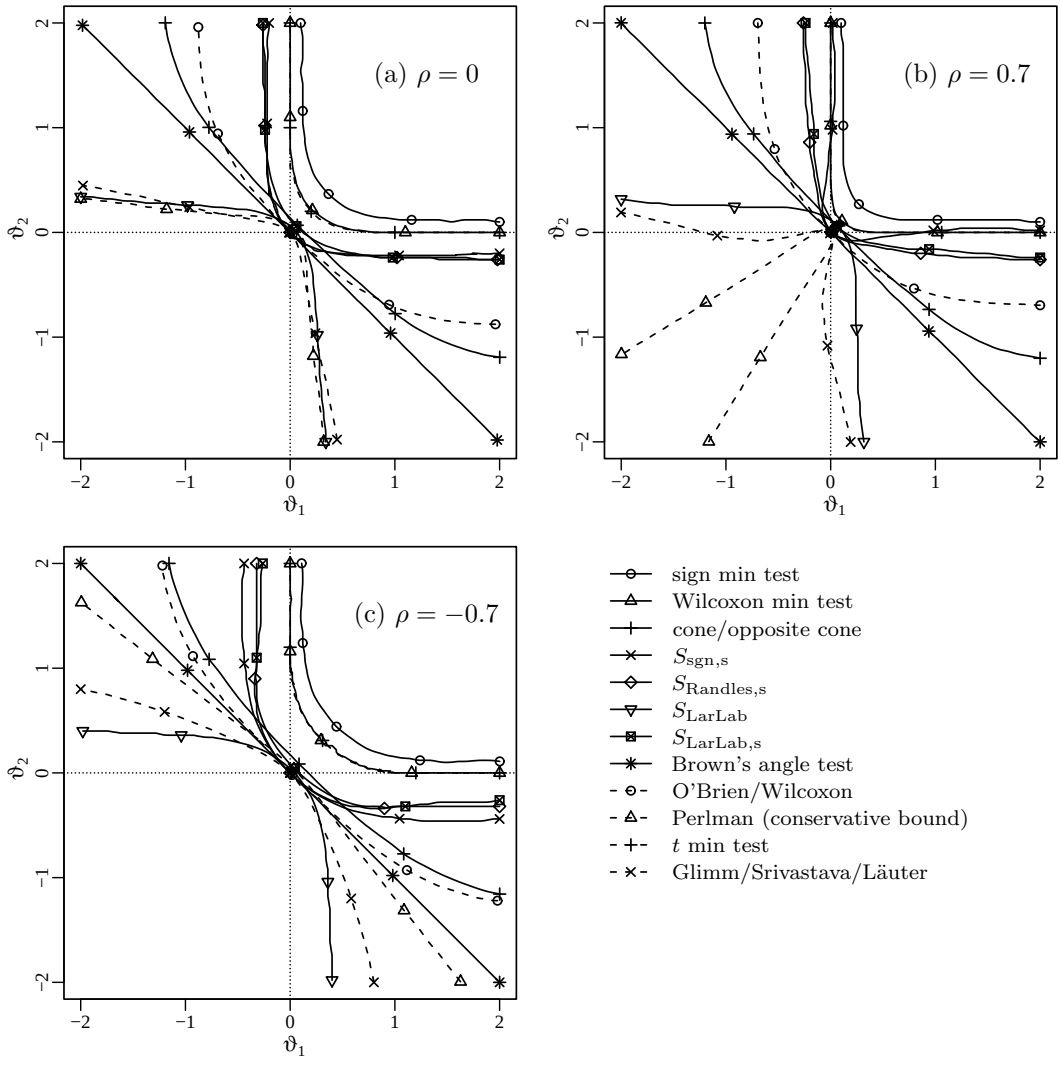

Figure 11.3: Approximate curves with power 0.05 for the tests with restricted alternatives under standard normal components with correlation $\rho \in \{0, 0.7, -0.7\}$. The symbols are used for identification purposes only; their positions on the curves have no particular meaning.

(e) *Liberal tests for a three-quadrant null hypothesis vs. a one-quadrant alternative*, potentially suitable for non-inferiority settings: Tests based on $S_{\text{sgn,s}}$, $S_{\text{Randles,s}}$, $S_{\text{LarLab,s}}$.

(f) *Conservative tests for a three-quadrant null hypothesis vs. a one-quadrant alternative*: sign min test, Wilcoxon min test, t min test.

(We have rearranged the classes such that the size of the null parameter region is in increasing order and that the alternatives become more and more specific.)

11.2 Effects of Correlation in the Normal Case

As a second step, we use a bivariate normal distribution with expectation $\boldsymbol{\vartheta}$ and covariance matrix $\left(\begin{smallmatrix} 1 & \rho \\ \rho & 1 \end{smallmatrix}\right)$, where $\rho \in \{-0.7, 0.7\}$.

We can see the effect of nonzero correlation on the curve of constant power in parts (b) and (c) of Figure 11.3. The curve of the test by Perlman (conservative bound) changes the most radically, especially with respect to the behavior for large $\|\boldsymbol{\vartheta}\|$; this test clearly does not respect the specified level for the one-quadrant null hypothesis in the positively correlated case. The test by Glimm, Srivastava, and Läuter, the O'Brien/Wilcoxon test, and the sector variant of the spatial sign test still show substantial changes in the shape of the curve. The curves of the other tests seem to be affected mainly near the origin, and these changes would not lead to a different judgement on the hypotheses that a test may be suitable for.

11.3 Heavy Tails: Bivariate $t(1)$

The noncentral bivariate t distribution with one degree of freedom serves us as an example of an elliptically symmetric distribution with heavy tails. The symmetry center $\boldsymbol{\vartheta}$ is the noncentrality parameter (median). The observations are generated as $\boldsymbol{Z}/\sqrt{Y} + \boldsymbol{\vartheta}$, where $\boldsymbol{Z}$ and Y are independent, Y has a χ^2 distribution with one degree of freedom, and $\boldsymbol{Z}$ is a bivariate normal random vector with mean $\mathbf{0}$, component variances 1, and correlation 0, 0.7, or -0.7.

Figure 11.4 shows that, when compared to the normal case, most of the curves do not change dramatically in shape. As can be expected, the three parametric tests are rather sensitive to the change in the underlying distribution. E. g., the t min test becomes less powerful than the sign min test. The most radical change concerns Perlman's test – even using the conservative bound, this test has a closed curve for $\rho = 0.7$, i. e. it has power above 0.05 for parameters on the main diagonal and below approximately $\boldsymbol{\vartheta} = (-2, -2)^{\mathrm{T}}$.

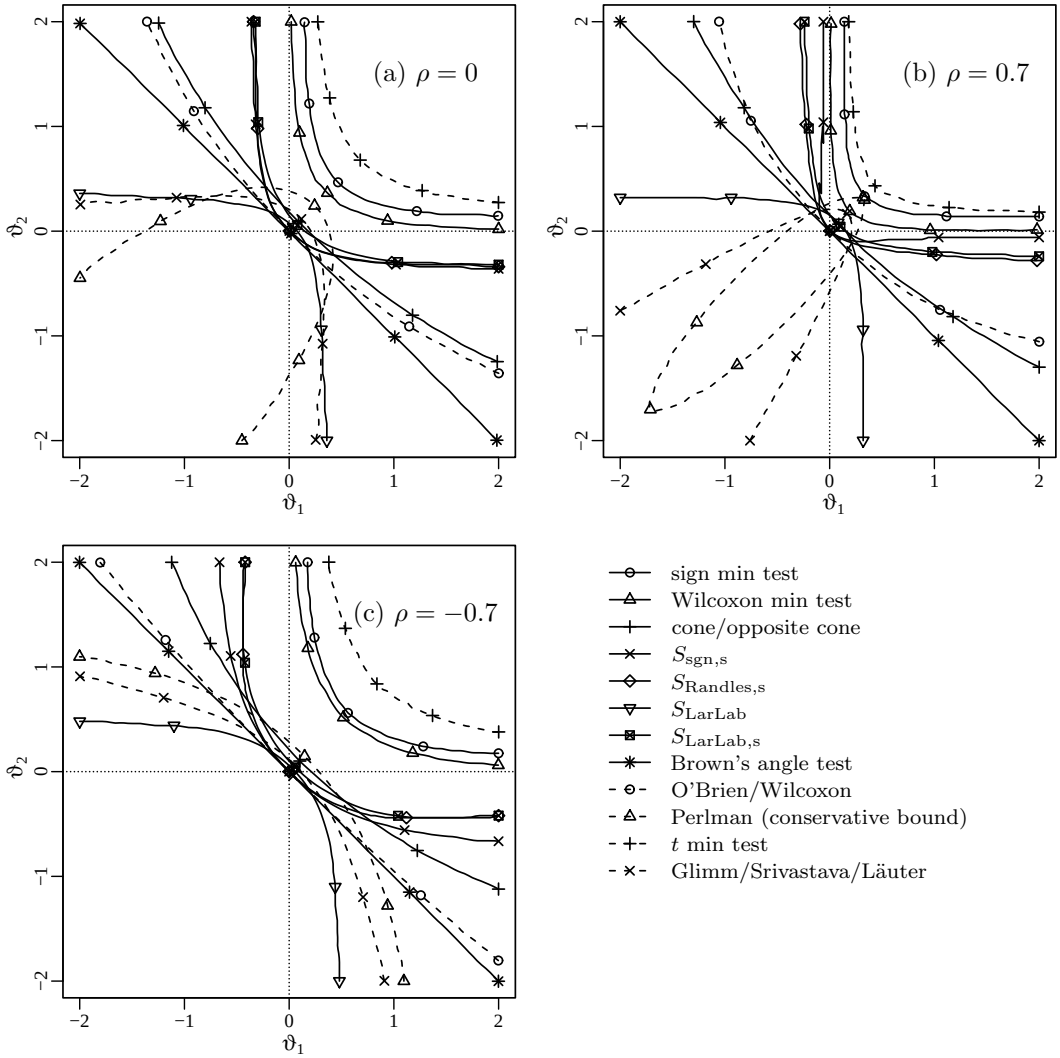

Figure 11.4: Approximate curves with power 0.05 for the tests with restricted alternatives under a noncentral bivariate $t(1)$ distribution, where the underlying bivariate normal distribution has correlation $\rho \in \{0, 0.7, -0.7\}$. The symbols are used for identification purposes only; their positions on the curves have no particular meaning.

11.4 "Exponential Radius" Distribution

The "exponential radius" distribution as defined in Larocque and Labarre (2004) is an example of a directionally (but not diagonally) symmetric distribution. It is generated by multiplying a unit vector having an angle distributed uniformly on $[0, \pi]$ by $\ln(2) - E$, the (possibly negative) "radius", where E has an exponential distribution with parameter 1 and is independent of the angle.

We can introduce dependence between the two components of the random vector by multiplication with the matrix

$$\begin{pmatrix} \sqrt{\frac{1}{2}\left(1+\sqrt{1-\rho^2}\right)} & \operatorname{sgn}\rho\sqrt{\frac{1}{2}\left(1-\sqrt{1-\rho^2}\right)} \\ \operatorname{sgn}\rho\sqrt{\frac{1}{2}\left(1-\sqrt{1-\rho^2}\right)} & \sqrt{\frac{1}{2}\left(1+\sqrt{1-\rho^2}\right)} \end{pmatrix},$$

which would transform a random vector with expectation $\mathbf{0}$ and covariance matrix $\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$ into one with covariance matrix $\left(\begin{smallmatrix} 1 & \rho \\ \rho & 1 \end{smallmatrix}\right)$. (Larocque and Labarre use a slightly different, but equivalent procedure for negative values of ρ.) In these simulations, we choose the values 0, 0.7, and -0.7 for ρ again. Finally, we add the location parameter $\boldsymbol{\vartheta}$.

In Figure 11.5, we see that the curves with power 0.05 are no longer symmetric with respect to the diagonal, but also that the shape of most of the curves does not change dramatically. The curves for the cone/opposite cone test and the O'Brien/Wilcoxon test now seem to be more similar to those of the tests for a three-quadrant null hypothesis against a one-quadrant alternative than in the previous cases. Perlman's test is again very sensitive to the dependency structure of the distribution.

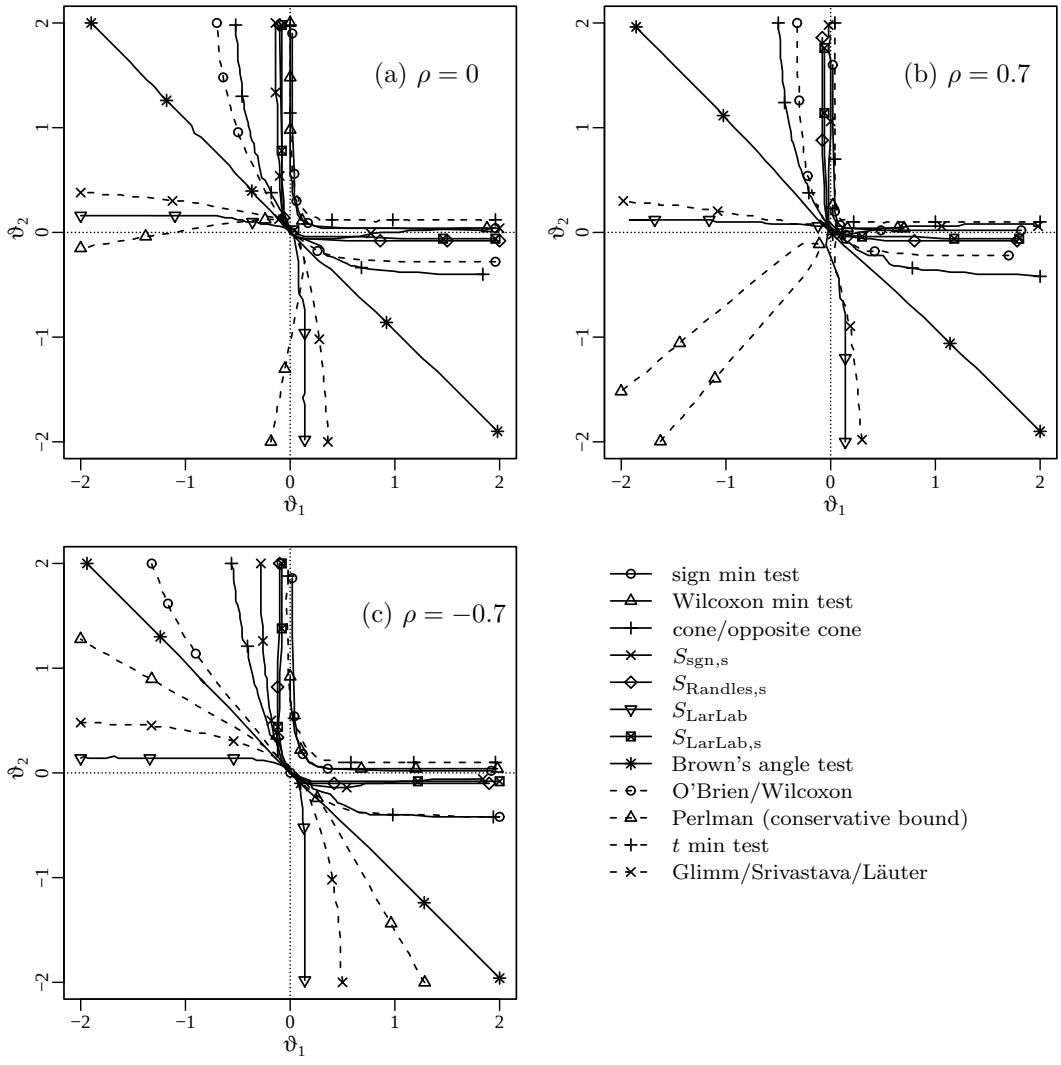

Figure 11.5: Approximate curves with power 0.05 for the tests with restricted alternatives under an exponential radius distribution, where the parameter ρ was chosen from $\{0, 0.7, -0.7\}$. The symbols are used for identification purposes only; their positions on the curves have no particular meaning.

Chapter 12

Confidence Regions

Statistical inference can be done by means of either hypothesis tests or confidence regions. It was observed by McDermott and Wang (1999) that there has not been much research about confidence regions for multivariate situations under order restrictions. Even in the parametric context, this observation still seems to hold. The purpose of this chapter is to translate some of the concepts formulated in terms of hypothesis tests for restricted alternatives into corresponding confidence regions and to obtain results about the shape of such confidence regions.

12.1 Connection with Hypothesis Tests

Since the connection between hypothesis tests and confidence regions is not as intuitive as in the univariate setting, it seems worthwhile to establish an accurate notational basis for confidence regions in a general parameter space. The definition of a confidence region that we use here is essentially taken from Witting (1985, pp. 289ff.), using an adapted notation. (Even though $\boldsymbol{\vartheta}$ may be from an arbitrary parameter space in this section, we use a bold face letter as we will apply the results in the context of location parameter vectors.)

Definition 12.1.1. Let $(\mathcal{X}, \mathcal{A}, (P_{\boldsymbol{\vartheta}})_{\boldsymbol{\vartheta}\in\Theta})$ be a probability space, i. e. $\mathcal{X}$ is the space that the observations occur in, $\mathcal{A}$ is a σ-algebra on $\mathcal{X}$, and $(P_{\boldsymbol{\vartheta}})_{\boldsymbol{\vartheta}\in\Theta}$ is a family of probability measures. Let $\gamma : \Theta \to \Gamma$ be a map into a suitable (for details, see the footnote in Witting, 1985, p. 24) measurable superset of $\gamma(\Theta)$, and $\alpha \in (0, 1)$. Further, for each $\boldsymbol{\vartheta} \in \Theta$, let $\Gamma_0(\boldsymbol{\vartheta}) \subset \Gamma$ be specified.

Based on the family $(\Gamma_0(\boldsymbol{\vartheta}))_{\boldsymbol{\vartheta}\in\Theta}$, a *confidence region* for $\gamma(\boldsymbol{\vartheta})$ with confidence level $1 - \alpha$ is a map $\mathcal{C}_{1-\alpha} : \mathcal{X} \to \mathcal{P}(\Gamma)$ such that

$$A(\boldsymbol{\gamma}) := \{x \in \mathcal{X} : \mathcal{C}_{1-\alpha}(x) \ni \boldsymbol{\gamma}\} \in \mathcal{A} \quad \forall\, \boldsymbol{\gamma} \in \Gamma$$

and

$$\mathrm{P}_{\boldsymbol{\vartheta}}(\mathcal{C}_{1-\alpha}(X) \ni \boldsymbol{\gamma}) \geq 1 - \alpha \quad \forall\, \boldsymbol{\gamma} \in \Gamma_0(\boldsymbol{\vartheta}) \quad \forall\, \boldsymbol{\vartheta} \in \Theta. \qquad \triangle$$

The first condition in Definition 12.1.1 only ensures measurability. Roughly speaking, the set $\Gamma_0(\boldsymbol{\vartheta})$ contains the values that should be in the confidence region if $\boldsymbol{\vartheta}$ is the true parameter.

There are simpler definitions (e. g. Borovkov, 1998, p. 285) of confidence regions that use $\gamma(\boldsymbol{\vartheta}) = \boldsymbol{\vartheta}$ and $\Gamma_0(\boldsymbol{\vartheta}) = \{\boldsymbol{\vartheta}\}, \forall\, \boldsymbol{\vartheta} \in \Theta$. The map $\gamma : \Theta \to \Gamma$ serves us mainly for notational clarity, i. e. for distinguishing between the parameter space Θ of the distribution and a meta-parameter space Γ; we will also normally use the canonical map $\gamma(\boldsymbol{\vartheta}) = \boldsymbol{\vartheta}$. The second simplification, $\Gamma_0(\boldsymbol{\vartheta}) = \{\boldsymbol{\vartheta}\}$, leads to confidence sets corresponding to tests based on simple null hypotheses only. We will use this simplification in Section 12.3.

Theorem 12.1.1. *Let* $(\mathcal{X}, \mathcal{A}, (P_{\boldsymbol{\vartheta}})_{\boldsymbol{\vartheta}\in\Theta})$, $\gamma : \Theta \to \Gamma$, α, $(\Gamma_0(\boldsymbol{\vartheta}))_{\boldsymbol{\vartheta}\in\Theta}$, *and* $(A(\boldsymbol{\gamma}))_{\boldsymbol{\gamma}\in\Gamma}$ *be as above. Define*

$$\Theta_0(\boldsymbol{\gamma}) := \{\boldsymbol{\vartheta} \in \Theta : \Gamma_0(\boldsymbol{\vartheta}) \ni \boldsymbol{\gamma}\}.$$

The following two statements are equivalent:

(a) $\mathcal{C}_{1-\alpha} : \mathcal{X} \to \mathcal{P}(\Gamma)$ *is a confidence region for* $\gamma(\boldsymbol{\vartheta})$ *with confidence level* $1-\alpha$ *based on the family* $(\Gamma_0(\boldsymbol{\vartheta}))_{\boldsymbol{\vartheta}\in\Theta}$.

(b) $\mathrm{P}_{\boldsymbol{\vartheta}}(X \in A(\boldsymbol{\gamma})) \geq 1-\alpha \quad \forall\, \boldsymbol{\vartheta} \in \Theta_0(\boldsymbol{\gamma}) \quad \forall\, \boldsymbol{\gamma} \in \Gamma.$

Proof.

(a)⇒(b) If $\mathcal{C}_{1-\alpha}$ is such a confidence region, then, by the definition of A,

$$\begin{aligned}\mathrm{P}_{\boldsymbol{\vartheta}}(X \in A(\boldsymbol{\gamma})) &= \mathrm{P}_{\boldsymbol{\vartheta}}(\mathcal{C}_{1-\alpha}(X) \ni \boldsymbol{\gamma}) \\ &\geq 1-\alpha \quad \forall\, \boldsymbol{\gamma} \in \Gamma_0(\boldsymbol{\vartheta}) \quad \forall\, \boldsymbol{\vartheta} \in \Theta.\end{aligned}$$

For arbitrary $\boldsymbol{\gamma} \in \Gamma$ and $\boldsymbol{\vartheta} \in \Theta_0(\boldsymbol{\gamma})$, $\boldsymbol{\gamma} \in \Gamma_0(\boldsymbol{\vartheta})$ holds by the definition of Θ_0, and (b) follows.

(b)⇒(a) The proof is completely analogous, using $\mathcal{C}_{1-\alpha}(x) = \{\boldsymbol{\gamma} : A(\boldsymbol{\gamma}) \ni x\}$ and $\Gamma_0(\boldsymbol{\vartheta}) = \{\boldsymbol{\gamma} : \Theta_0(\boldsymbol{\gamma}) \ni \boldsymbol{\vartheta}\}$. □

According to statement (b) in Theorem 12.1.1, for each $\boldsymbol{\gamma} \in \Gamma$, $A(\boldsymbol{\gamma})$ is the acceptance region of a non-randomized level α test of $H_0 : \boldsymbol{\vartheta} \in \Theta_0(\boldsymbol{\gamma})$. Therefore, the desired correspondence between hypothesis tests and confidence regions is established.

Definition 12.1.2. Using the notation from Definition 12.1.1, let $\mathcal{C}_{1-\alpha}$ be a confidence region for $\gamma(\boldsymbol{\vartheta})$. Let $\Gamma_1(\boldsymbol{\vartheta}) \subset \Gamma \setminus \Gamma_0(\boldsymbol{\vartheta}), \forall\, \boldsymbol{\vartheta} \in \Theta$.

$\mathcal{C}_{1-\alpha}$ is *unbiased* with respect to $(\Gamma_1(\boldsymbol{\vartheta}))_{\boldsymbol{\vartheta}\in\Theta}$ if

$$\mathrm{P}_{\boldsymbol{\vartheta}}(\mathcal{C}_{1-\alpha}(X) \ni \boldsymbol{\gamma}) \leq 1-\alpha \quad \forall\, \boldsymbol{\gamma} \in \Gamma_1(\boldsymbol{\vartheta}) \quad \forall\, \boldsymbol{\vartheta} \in \Theta. \qquad \triangle$$

Theorem 12.1.2. *Let* $(\mathcal{X}, \mathcal{A}, (P_{\boldsymbol{\vartheta}})_{\boldsymbol{\vartheta}\in\Theta})$, $\gamma : \Theta \to \Gamma$, α, $(\Gamma_0(\boldsymbol{\vartheta}))_{\boldsymbol{\vartheta}\in\Theta}$, $(\Gamma_1(\boldsymbol{\vartheta}))_{\boldsymbol{\vartheta}\in\Theta}$, $(A(\boldsymbol{\gamma}))_{\boldsymbol{\gamma}\in\Gamma}$, *and* $(\Theta_0(\boldsymbol{\gamma}))_{\boldsymbol{\gamma}\in\Gamma}$ *be as above. Define*

$$\Theta_1(\boldsymbol{\gamma}) := \{\boldsymbol{\vartheta} \in \Theta : \Gamma_1(\boldsymbol{\vartheta}) \ni \boldsymbol{\gamma}\}.$$

Assume that $\mathcal{C}_{1-\alpha} : \mathcal{X} \to \mathcal{P}(\Gamma)$ *is a confidence region for* $\gamma(\boldsymbol{\vartheta})$ *with confidence level* $1-\alpha$ *based on the family* $(\Gamma_0(\boldsymbol{\vartheta}))_{\boldsymbol{\vartheta}\in\Theta}$.

The following two statements are equivalent:

(a) $\mathcal{C}_{1-\alpha}$ *is unbiased with respect to* $(\Gamma_1(\boldsymbol{\vartheta}))_{\boldsymbol{\vartheta}\in\Theta}$.

(b) $\mathrm{P}_{\boldsymbol{\vartheta}}(X \in A(\boldsymbol{\gamma})) \leq 1-\alpha \quad \forall\, \boldsymbol{\vartheta} \in \Theta_1(\boldsymbol{\gamma}) \quad \forall\, \boldsymbol{\gamma} \in \Gamma.$

Proof. Completely analogous to the proof of Theorem 12.1.1. □

Note that $\Gamma_1(\boldsymbol{\vartheta}) \subset \Gamma \setminus \Gamma_0(\boldsymbol{\vartheta})$, $\forall\, \boldsymbol{\vartheta} \in \Theta$, is equivalent to $\Theta_1(\boldsymbol{\gamma}) \subset \Theta \setminus \Theta_0(\boldsymbol{\gamma})$, $\forall\, \boldsymbol{\gamma} \in \Gamma$.

Theorems 12.1.1 and 12.1.2 together state the equivalence of unbiased non-randomized level α tests and unbiased confidence regions with confidence level $1-\alpha$. If we consider tests for sector or cone alternatives about a multivariate location parameter with composite null hypotheses (i. e. $\Theta_1(\boldsymbol{\gamma}) = \mathbb{R}^2 \setminus \Theta_0(\boldsymbol{\gamma}), \forall\, \boldsymbol{\gamma} \in \Gamma$), this equivalence combined with the results from Section 3.2 means that, in general, we cannot expect to obtain unbiased confidence regions by inverting such tests.

Equivalences of uniformly most powerful (unbiased) tests and confidence regions can also be established – see Witting (1985). Since our focus is primarily on the appropriateness of methods in terms of respecting the level (and not that much on optimal power), we do not consider these optimality properties here.

12.2 Shape of the Confidence Region for the Meta-Parameter $\boldsymbol{\gamma}$

Theorem 12.1.1 implies that for each non-randomized test about some parameter vector $\boldsymbol{\vartheta}$, a corresponding confidence region for the meta-parameter $\gamma(\boldsymbol{\vartheta})$ according to Definition 12.1.1 exists. We can use the connection with the acceptance region of the test to characterize the shape of the confidence region.

12.2.1 Translations of the Acceptance Region in $\mathbb{R}^p$

One of the simplest cases is the following one, where the parameter $\boldsymbol{\vartheta}$, the observations, and the test statistic are all of the same dimension, $\Gamma = \Theta$, $\gamma(\boldsymbol{\vartheta}) = \boldsymbol{\vartheta}$, $\forall\, \boldsymbol{\vartheta} \in \Theta$, and the decision rule is particularly simple:

Theorem 12.2.1. *Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be i. i. d. random vectors in $\mathbb{R}^p$ having a distribution $F_{\boldsymbol{\vartheta}}$, with $\boldsymbol{\vartheta} \in \Theta \subset \mathbb{R}^p$. Let a non-randomized level α test for $H_0 : \boldsymbol{\vartheta} \in \Theta_0(\boldsymbol{\gamma})$ vs. $H_1 : \boldsymbol{\vartheta} \in \Theta \setminus \Theta_0(\boldsymbol{\gamma})$ be given, and assume that, for each $\boldsymbol{\gamma} \in \Theta$, this test accepts H_0 if and only if some test statistic $T(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n) \in \mathbb{R}^p$ is within the acceptance region $A(\boldsymbol{\gamma})$.*

If the acceptance region is of the form

$$A(\boldsymbol{\gamma}) = \boldsymbol{\gamma} + A$$

for some $A \subset \mathbb{R}^p$, then a confidence region for $\boldsymbol{\gamma}$ with confidence level $1 - \alpha$ is given by

$$\mathcal{C}_{1-\alpha}(T(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)) = T(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n) - A.$$

Proof. According to Theorem 12.1.1, the confidence region corresponding to the given test is

$$\begin{aligned}
\mathcal{C}_{1-\alpha}(T(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)) &= \{\boldsymbol{\gamma} : A(\boldsymbol{\gamma}) \ni T(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)\} \\
&= \{\boldsymbol{\gamma} : \boldsymbol{\gamma} + A \ni T(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)\} \\
&= \{\boldsymbol{\gamma} : \boldsymbol{\gamma} \in T(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n) - A\} \\
&= T(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n) - A.
\end{aligned}$$

□

The assumptions in Theorem 12.2.1 are rather restrictive. However, we can use it for a formal proof that the confidence region corresponding to a min test (as defined in Section 5.2) consists of all points that are above the (univariate) lower confidence bound with respect to at least one component:

Corollary 12.2.2. *Let the $1 - \alpha$ lower confidence bounds for γ_i corresponding to univariate tests for $H_{0i} : \vartheta_i \leq \gamma_i$ vs. $H_{1i} : \vartheta_i > \gamma_i$ be given by $t_i(X_{1i}, \ldots, X_{ni})$. Let a min test for $H_0 : \exists\, i \in \{1, \ldots, p\} : \vartheta_i \leq \gamma_i$ vs. $H_1 : \boldsymbol{\vartheta} > \boldsymbol{\gamma}$ be based on these univariate tests, i. e. it rejects H_0 at level α if each univariate tests rejects its H_{0i} at level α.*

The $1 - \alpha$ confidence region for $\boldsymbol{\gamma}$ corresponding to this min test can then be written as

$$(t_1(X_{11}, \ldots, X_{n1}), \ldots, t_p(X_{1p}, \ldots, X_{np}))^{\mathrm{T}} + (\mathbb{R}^p \setminus \mathbb{R}^p_-).$$

Proof. Define $T(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n) := (t_1(X_{11}, \ldots, X_{n1}), \ldots, t_p(X_{1p}, \ldots, X_{np}))^{\mathrm{T}}$. H_0 is rejected if and only if $\gamma_i < t_i(X_{1i}, \ldots, X_{ni}), \forall\, i \in \{1, \ldots, p\}$, which is equivalent to $T(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n) \in \boldsymbol{\gamma} + \mathbb{R}^p_+$. The acceptance region for H_0 is therefore $A(\boldsymbol{\gamma}) = \mathbb{R}^p \setminus (\boldsymbol{\gamma} + \mathbb{R}^p_+) = \boldsymbol{\gamma} + (\mathbb{R}^p \setminus \mathbb{R}^p_+)$, such that we can apply Theorem 12.2.1 with $A = \mathbb{R}^p \setminus \mathbb{R}^p_+$, yielding that the $1 - \alpha$ confidence region for $\boldsymbol{\gamma}$ is

$$T(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n) - (\mathbb{R}^p \setminus \mathbb{R}^p_+) = T(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n) + (\mathbb{R}^p \setminus \mathbb{R}^p_-).$$

□

This result may be astonishing at first glance: While a min test can be used to show that $\boldsymbol{\vartheta} \in \mathbb{R}^2$ is in the first quadrant, the corresponding confidence region is a translated version of the first, second, and fourth quadrant, and it therefore always contains parts of the second and fourth quadrant. However, we should remember that our confidence regions are for the meta-parameter $\boldsymbol{\gamma}$ and that excluding $\boldsymbol{\gamma}$ from the confidence region corresponds to rejecting the null hypothesis that $\boldsymbol{\vartheta} \in \Theta_0(\boldsymbol{\gamma}) = \boldsymbol{\gamma} + (\mathbb{R}^2 \setminus \mathbb{R}^2_+)$; it is evident that we cannot reject this null hypothesis with a sensible min test if e. g. γ_1 is above every X_{i1}, even if γ_2 is very low.

We will investigate in Sections 12.3 and 12.4 how confidence regions for $\boldsymbol{\vartheta}$ itself can be obtained. For the time being, we try to obtain a result that is similar to the one above under more general conditions.

12.2.2 Bounds for the Border of the Confidence Region

For a min test, we can write the confidence region as $T(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n) - \Theta_0(\mathbf{0})$. Even in more general situations, a similar relationship seems to hold. However, the shape of the confidence region often only corresponds to the shape given above in an asymptotic sense, i. e. for parameter values that are distant enough from the observations.

In order to measure distances between sets, we need the following definitions:

Definition 12.2.1. Let M be a metric space (with metric d) with an origin $\mathbf{0}$, and let B_r be the open ball of radius r, i. e. $B_r = \{\boldsymbol{x} \in M : \mathrm{d}(\mathbf{0}, \boldsymbol{x}) < r\}$. Let A_1 and A_2 be subsets of M.

(a) The *infimum distance* between A_1 and A_2 is

$$\mathrm{d}_{\inf}(A_1, A_2) = \inf_{\boldsymbol{a}_1 \in A_1, \boldsymbol{a}_2 \in A_2} \mathrm{d}(\boldsymbol{a}_1, \boldsymbol{a}_2).$$

(b) The *Hausdorff distance* between A_1 and A_2 is

$$\mathrm{d}_{\mathrm{H}}(A_1, A_2) = \max\{ \sup_{\boldsymbol{a}_1 \in A_1} \inf_{\boldsymbol{a}_2 \in A_2} \mathrm{d}(\boldsymbol{a}_1, \boldsymbol{a}_2), \sup_{\boldsymbol{a}_2 \in A_2} \inf_{\boldsymbol{a}_1 \in A_1} \mathrm{d}(\boldsymbol{a}_1, \boldsymbol{a}_2)\}.$$

(c) The *punched Hausdorff distance with radius r* between A_1 and A_2 is

$$\mathrm{d}^r_{\mathrm{H}}(A_1, A_2) = \max\{ \sup_{\boldsymbol{a}_1 \in A_1 \cap B_r^{\mathrm{c}}} \inf_{\boldsymbol{a}_2 \in A_2} \mathrm{d}(\boldsymbol{a}_1, \boldsymbol{a}_2), \sup_{\boldsymbol{a}_2 \in A_2 \cap B_r^{\mathrm{c}}} \inf_{\boldsymbol{a}_1 \in A_1} \mathrm{d}(\boldsymbol{a}_1, \boldsymbol{a}_2)\}. \quad \triangle$$

The infimum distance is not a metric (it lacks positivity for overlapping, but non-identical sets).

The Hausdorff distance is only a metric if we restrict attention to the family of compact non-empty subsets of M. However, for the comparison of confidence regions, the difference between sets with the same closure does not seem to be

of much importance, such that we can use this distance measure for any two non-empty sets – even for unbounded sets if we are willing to accept distances of ∞.

We introduce the punched Hausdorff distance here to measure how well two sets agree for points far away from the origin, i. e. after "punching out" the area around the origin. Therefore, it inevitably lacks positivity when looking at sets that differ only near the origin. The advantage of looking at this definition instead of just $\mathrm{d_H}(A_1 \cap B_r^{\mathrm{c}}, A_2 \cap B_r^{\mathrm{c}})$ is that some unpleasant properties can be avoided; especially, $\mathrm{d_H^r}(A_1, A_2) \leq \mathrm{d_H}(A_1, A_2)$ holds true for every $0 \leq r < \infty$.

The infimum distance corresponds to the perception of the distance between two objects in everyday life and is 0 whenever two sets (or their closures) have any point in common. The Hausdorff distance, in contrast, is only 0 if the closures of both sets agree completely.

We assume in the following that $\Theta = \Gamma = \mathbb{R}^p$. As a first step towards a more general result about the shape of the confidence region, Lemma 12.2.3 gives conditions ensuring that the confidence region contains some set of the supposed shape. (In the following two lemmas, the assumptions are perhaps most plausible and easier to understand when C is taken to be the alternative parameter region $\mathbb{R}^p \setminus \Theta_0(\mathbf{0})$, but the more general versions given here will be useful.)

Lemma 12.2.3. *Let a non-randomized level α test for $H_0 : \boldsymbol{\vartheta} \in \Theta_0(\boldsymbol{\gamma})$ vs. $H_1 : \boldsymbol{\vartheta} \in \mathbb{R}^p \setminus \Theta_0(\boldsymbol{\gamma})$ be given, where $\Theta_0(\boldsymbol{\gamma}) = \boldsymbol{\gamma} + \Theta_0(\mathbf{0}) \subset \mathbb{R}^p$, $\forall\, \boldsymbol{\gamma} \in \mathbb{R}^p$, the test being based on random vectors $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n \in \mathbb{R}^p$.*

Let $C \subset \mathbb{R}^p$, $C \neq \mathbb{R}^p$, be a non-empty cone. Assume that

(a) *C is convex or*

(b) *$\mathbb{R}^p \setminus C$ is convex,*

*and that for each set of observed vectors $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$, there is a $d_0 < \infty$ such that $\mathrm{d_{inf}}(\mathrm{conv}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n), \boldsymbol{\gamma} + C) > d_0$ implies that $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) \in A(\boldsymbol{\gamma})$, the acceptance region of the test. (*conv *denotes the convex hull.)*

Then there exists a point $\boldsymbol{c}^*(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) \in \mathbb{R}^p$ such that $\boldsymbol{c}^*(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) - (\mathbb{R}^p \setminus C)$ is entirely within the $1 - \alpha$ confidence region $\mathcal{C}_{1-\alpha}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ corresponding to the test.

Proof.

(a) $\mathrm{conv}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ is bounded, and so is $\mathrm{conv}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) + B_{d_0+1}$, with $B_d = \{\boldsymbol{x} \in \mathbb{R}^p : \|\boldsymbol{x}\| < d\}$. $\mathrm{conv}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) + B_{d_0+1} + (\mathbb{R}^p \setminus C)$ is therefore a true subset of $\mathbb{R}^p$, such that we can choose a $\boldsymbol{c}^*(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ in its complement.

Let $\boldsymbol{\gamma} = \boldsymbol{c}^*(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) - \boldsymbol{\vartheta}^*$ with arbitrary $\boldsymbol{\vartheta}^* \in \mathbb{R}^p \setminus C$. Then

$$\begin{aligned} \boldsymbol{\gamma} + C &= \boldsymbol{c}^*(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) - \boldsymbol{\vartheta}^* + C \\ &\subset \boldsymbol{c}^*(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) - (\mathbb{R}^p \setminus C) \end{aligned}$$

due to Lemma B.5.1. Therefore,

$$\begin{aligned}
&\mathrm{d}_{\inf}\left(\operatorname{conv}(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n),\boldsymbol{\gamma}+C\right)\\
&\quad\geq \mathrm{d}_{\inf}\left(\operatorname{conv}(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n),\boldsymbol{c}^*(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n)-(\mathbb{R}^p\setminus C)\right)\\
&\quad= \mathrm{d}_{\inf}\left(\operatorname{conv}(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n)+(\mathbb{R}^p\setminus C),\{\boldsymbol{c}^*(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n)\}\right)\\
&\quad\geq d_0+1>d_0,
\end{aligned}$$

the second inequality following from the choice of $\boldsymbol{c}^*(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n)$. Because $\boldsymbol{\vartheta}^*\in\mathbb{R}^p\setminus C$ was arbitrary and by the assumption made, this implies that

$$(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n)\in A(\boldsymbol{\gamma})\quad\forall\,\boldsymbol{\gamma}\in\boldsymbol{c}^*(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n)-(\mathbb{R}^p\setminus C),$$

and therefore

$$\boldsymbol{c}^*(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n)-(\mathbb{R}^p\setminus C)\subset\mathcal{C}_{1-\alpha}(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n).$$

(b) Only a few changes are necessary to adapt the proof of (a):

Choose $\boldsymbol{c}^*(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n)$ in the complement of $\operatorname{conv}(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n)+B_{d_0+1}-C$. With $\boldsymbol{\vartheta}^*\in\mathbb{R}^p\setminus C$ and $\boldsymbol{\gamma}=\boldsymbol{c}^*(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n)-\boldsymbol{\vartheta}^*$, use Lemma B.5.1 again to see that

$$\begin{aligned}
\boldsymbol{\gamma}+C&=\boldsymbol{c}^*(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n)-\boldsymbol{\vartheta}^*+C\\
&\subset\boldsymbol{c}^*(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n)+C.
\end{aligned}$$

It then follows that

$$\begin{aligned}
&\mathrm{d}_{\inf}\left(\operatorname{conv}(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n),\boldsymbol{\gamma}+C\right)\\
&\quad\geq \mathrm{d}_{\inf}\left(\operatorname{conv}(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n),\boldsymbol{c}^*(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n)+C\right)\\
&\quad= \mathrm{d}_{\inf}\left(\operatorname{conv}(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n)-C,\{\boldsymbol{c}^*(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n)\}\right)\\
&\quad\geq d_0+1>d_0,
\end{aligned}$$

and the rest of the proof can be taken from the proof of (a) again. □

In order to approach the border of the confidence region from the side of its complement, we use the cone order monotonicity property of tests defined in Section 3.3:

Lemma 12.2.4. *Let a non-randomized level α test for $H_0:\ \boldsymbol{\vartheta}\in\Theta_0(\boldsymbol{\gamma})$ vs. $H_1:\ \boldsymbol{\vartheta}\in\mathbb{R}^p\setminus\Theta_0(\boldsymbol{\gamma})$ be given, where $\Theta_0(\boldsymbol{\gamma})=\boldsymbol{\gamma}+\Theta_0(\mathbf{0})\subset\mathbb{R}^p$, $\forall\,\boldsymbol{\gamma}\in\mathbb{R}^p$, the test being based on random vectors $\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n\in\mathbb{R}^p$. Let the test be translation invariant and cone order monotone in the sample with respect to the convex cone C.*

Then the $1-\alpha$ confidence region $\mathcal{C}_{1-\alpha}(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n)$ corresponding to the test has the following properties:

(a) $\boldsymbol{\gamma} \in \mathcal{C}_{1-\alpha}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) \Rightarrow \boldsymbol{\gamma} + C \subset \mathcal{C}_{1-\alpha}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$

(b) $\boldsymbol{\gamma} \notin \mathcal{C}_{1-\alpha}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) \Rightarrow \boldsymbol{\gamma} - C \subset \mathbb{R}^p \setminus \mathcal{C}_{1-\alpha}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$

Proof. Statement (b) is more important for us, and we therefore prove it first.

(b) The required cone order monotonicity in the sample for a non-randomized test φ means that, for all $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \in \mathbb{R}^p, \boldsymbol{\gamma} \in \mathbb{R}^p, \boldsymbol{\delta} \in C$,

$$\varphi((\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n), \boldsymbol{\gamma}) = 1 \Rightarrow \varphi((\boldsymbol{x}_1 + \boldsymbol{\delta}, \ldots, \boldsymbol{x}_n + \boldsymbol{\delta}), \boldsymbol{\gamma}) = 1, \qquad (*)$$

where $\varphi((\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n), \boldsymbol{\gamma}) = 1$ denotes the rejection of $H_0 : \boldsymbol{\vartheta} \in \Theta_0(\boldsymbol{\gamma})$.

Take any $\boldsymbol{\gamma} \notin \mathcal{C}_{1-\alpha}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ and $\boldsymbol{\delta} \in C$. Using $(*)$ and the translation invariance (t. i.), we obtain

$$\begin{aligned}
\boldsymbol{\gamma} \notin \mathcal{C}_{1-\alpha}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) &\Leftrightarrow (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) \notin A(\boldsymbol{\gamma}) \\
&\Leftrightarrow \varphi((\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n), \boldsymbol{\gamma}) = 1 \\
&\overset{(*)}{\Rightarrow} \varphi((\boldsymbol{x}_1 + \boldsymbol{\delta}, \ldots, \boldsymbol{x}_n + \boldsymbol{\delta}), \boldsymbol{\gamma}) = 1 \\
&\overset{\text{t. i.}}{\Leftrightarrow} \varphi((\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n), \boldsymbol{\gamma} - \boldsymbol{\delta}) = 1 \\
&\Leftrightarrow (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) \notin A(\boldsymbol{\gamma} - \boldsymbol{\delta}) \\
&\Leftrightarrow \boldsymbol{\gamma} - \boldsymbol{\delta} \notin \mathcal{C}_{1-\alpha}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n),
\end{aligned}$$

which completes the proof of (b).

(a) Let $\boldsymbol{\gamma} \in \mathbb{R}^p$, $\boldsymbol{\delta} \in C$. From (b),

$$\begin{aligned}
& \boldsymbol{\gamma} + \boldsymbol{\delta} \notin \mathcal{C}_{1-\alpha}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) \Rightarrow \boldsymbol{\gamma} \notin \mathcal{C}_{1-\alpha}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) \\
\Leftrightarrow \quad & \boldsymbol{\gamma} + \boldsymbol{\delta} \in \mathcal{C}_{1-\alpha}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) \Leftarrow \boldsymbol{\gamma} \in \mathcal{C}_{1-\alpha}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n). \qquad \square
\end{aligned}$$

12.2.3 Bivariate Case

In the bivariate case, i. e. $\Theta = \Gamma = \mathbb{R}^2$, the combination of the two lemmas above gives us enough information to determine the asymptotic (for $\|\boldsymbol{\gamma}\| \to \infty$) shape of the confidence region in the case where C (or again, as a special case, $\mathbb{R}^2 \setminus \Theta_0(\mathbf{0})$) is a convex sector.

Theorem 12.2.5. *Let a non-randomized level α test for $H_0 : \boldsymbol{\vartheta} \in \Theta_0(\boldsymbol{\gamma})$ vs. $H_1 : \boldsymbol{\vartheta} \in \mathbb{R}^2 \setminus \Theta_0(\boldsymbol{\gamma})$ be given, where $\Theta_0(\boldsymbol{\gamma}) = \boldsymbol{\gamma} + \Theta_0(\mathbf{0}) \subset \mathbb{R}^2$, $\forall\, \boldsymbol{\gamma} \in \mathbb{R}^2$, the test being based on random vectors $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n \in \mathbb{R}^2$. Let the test be translation invariant and cone order monotone in the sample with respect to the non-empty sector $C \subset \mathbb{R}^2$ with an angle strictly less than π. Assume that for each set of observed vectors $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$, there is a $d_0 < \infty$ such that $\mathrm{d}_{\inf}(\mathrm{conv}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n), \boldsymbol{\gamma} + C) > d_0$ implies that $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) \in A(\boldsymbol{\gamma})$, the acceptance region of the test. Assume*

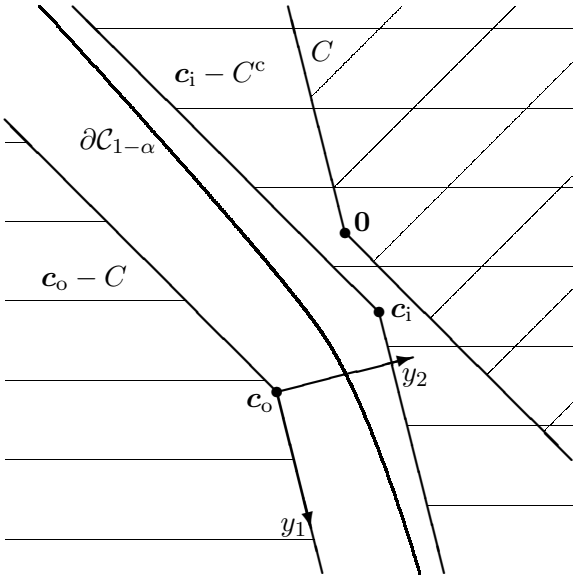

Figure 12.1: Proof of Theorem 12.2.5: Application of Lemmas 12.2.3 (a) and 12.2.4 (b) to obtain an angled band that must contain $\partial\mathcal{C}_{1-\alpha}$. $C^\mathrm{c} = \mathbb{R}^2 \setminus C$.

further that the $1-\alpha$ *confidence region* $\mathcal{C}_{1-\alpha}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ *corresponding to the test is a true subset of* $\mathbb{R}^2$.

Then there exists a $c^*(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) \in \mathbb{R}^2$ *such that*

$$\mathrm{d}_\mathrm{H}^r(\mathcal{C}_{1-\alpha}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n), c^*(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) - (\mathbb{R}^2 \setminus C)) \to 0 \quad (r \to \infty).$$

Proof. According to Lemma 12.2.3 (a), there exists a point $\boldsymbol{c}_\mathrm{i}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) \in \mathbb{R}^2$ such that $\boldsymbol{c}_\mathrm{i}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) - (\mathbb{R}^2 \setminus C)$ is entirely within $\mathcal{C}_{1-\alpha} = \mathcal{C}_{1-\alpha}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ ("inner bound" for $\partial\mathcal{C}_{1-\alpha}$). We take any point $\boldsymbol{c}_\mathrm{o}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) \notin \mathcal{C}_{1-\alpha}$ (which exists by assumption) and obtain from Lemma 12.2.4 (b) that $\boldsymbol{c}_\mathrm{o}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) - C$ is completely outside $\mathcal{C}_{1-\alpha}$ ("outer bound" for $\partial\mathcal{C}_{1-\alpha}$). Therefore, $\partial\mathcal{C}_{1-\alpha}$ has to lie within an angled band, as is illustrated for an example in Figure 12.1.

Take now any positive ε.

Considering in a first step only one part of the angled band (e. g. the lower right part in Figure 12.1), we can use a new coordinate system (y_1, y_2) as indicated in the figure. Let $y_1(\boldsymbol{x})$ and $y_2(\boldsymbol{x})$ denote the coordinates of $\boldsymbol{x}$ in this new coordinate system.

We can interpret $\{(y_1(\boldsymbol{x}), y_2(\boldsymbol{x})) : \boldsymbol{x} \in \partial\mathcal{C}_{1-\alpha}\}$ as the graph of a function $f : y_1 \mapsto y_2$ – more precisely, we take $f(y_1) = \sup\{y_2 : y_1 = y_1(\boldsymbol{x}), y_2 = y_2(\boldsymbol{x})$ for some $\boldsymbol{x} \in \partial\mathcal{C}_{1-\alpha}\}$, which does not make any difference asymptotically due to the monotonicity ensured by Lemma 12.2.4 (b). $f(y_1)$ is monotone and bounded and therefore converges to a limit f_∞ for $y_1 \to \infty$, i. e. $\partial\mathcal{C}_{1-\alpha}$ is asymptotically parallel to $\partial(\boldsymbol{c}_\mathrm{i} - (\mathbb{R}^2 \setminus C))$. Therefore, there exists $\tilde{r}_1 < \infty$ such that $f(y_1(\boldsymbol{x})) \in (f_\infty - \varepsilon, f_\infty]$ for $\|\boldsymbol{x}\| > \tilde{r}_1$. We increase this radius $\tilde{r}_1$ by the width of the band

and then call the increased radius r_1. This ensures that also each point $\boldsymbol{x}$ on the asymptote of $\partial\mathcal{C}_{1-\alpha}$ with $\|\boldsymbol{x}\| > r_1$ is at most ε away from $\partial\mathcal{C}_{1-\alpha}$.

In the same way, we define a radius r_2 for the other part of the angled band that must contain $\partial\mathcal{C}_{1-\alpha}$, and we get the second asymptote of $\partial\mathcal{C}_{1-\alpha}$. Define $r = \max\{r_1, r_2\}$, and choose $c^*(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ as the intersection of both asymptotes of $\partial\mathcal{C}_{1-\alpha}$ (which is possible since C is not a half-plane, and therefore, the asymptotes are not parallel). It follows by the construction of r that $\mathrm{d}_{\mathrm{H}}^r(\mathcal{C}_{1-\alpha}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n), c^*(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) - (\mathbb{R}^2 \setminus C)) < \varepsilon$. □

We have used a general sector C (instead of $\mathbb{R}^2 \setminus \Theta_0(\mathbf{0})$, which would be most intuitive) in Theorem 12.2.5. This has the advantage that we can also apply the theorem to tests without knowing the (largest) composite null hypothesis that they can be used for, as can be seen from the following example:

Example 12.2.1. We use the cone/opposite cone test (Section 6.2) for a sector C with an angle less than π. This test does not respect the level α for the composite null hypothesis that $\boldsymbol{\vartheta} \in \boldsymbol{\gamma} + (\mathbb{R}^2 \setminus C)$, and the composite null hypothesis that obviously ensures the level to be respected, $\boldsymbol{\vartheta} \in \boldsymbol{\gamma} - C$, is quite far away from the shape suggested by the plots in Chapter 11.

However, we know that the test is translation invariant and cone order monotone with respect to C. Further, it is easy to see that $\mathrm{d}_{\inf}(\mathrm{conv}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n), \boldsymbol{\gamma} + C) > 0$ implies that $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) \in A(\boldsymbol{\gamma})$ – the test will not be able to reject at any level $\alpha < 0.5$ if all observations are outside of $\boldsymbol{\gamma} + C$. Whenever the $1-\alpha$ confidence region obtained by inverting this test is not the entire parameter space $\mathbb{R}^2$, Theorem 12.2.5 can be applied, and the confidence region is therefore asymptotically (for $\|\boldsymbol{\gamma}\| \to \infty$) a translated version of $-(\mathbb{R}^2 \setminus C)$. △

12.3 Confidence Regions for $\boldsymbol{\vartheta}$ Based on Simple Null Hypotheses

From Definition 12.1.1, we obtain confidence regions for $\boldsymbol{\gamma} = \gamma(\boldsymbol{\vartheta})$, the meta-parameter that defines the null and alternative parameter regions $\Theta_0(\boldsymbol{\gamma})$ and $\Theta_1(\boldsymbol{\gamma})$. The conditions of the following theorem ensure that we can use the same confidence regions for $\boldsymbol{\vartheta}$ itself. (We return to the general notation and write X for the observed data again, instead of $(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)$.)

Theorem 12.3.1. *With the notation as in Definition 12.1.1, let $\mathcal{C}_{1-\alpha} : \mathcal{X} \to \mathcal{P}(\Gamma)$ be a confidence region for $\gamma(\boldsymbol{\vartheta})$ with confidence level $1-\alpha$ based on the family $(\Gamma_0(\boldsymbol{\vartheta}))_{\boldsymbol{\vartheta}\in\Theta}$.*

If $\Theta = \Gamma$, $\gamma(\boldsymbol{\vartheta}) = \boldsymbol{\vartheta}, \forall\, \boldsymbol{\vartheta} \in \Theta$, and $\boldsymbol{\vartheta} \in \Gamma_0(\boldsymbol{\vartheta}), \forall\, \boldsymbol{\vartheta} \in \Theta$, then

$$\mathrm{P}_{\boldsymbol{\vartheta}}(\mathcal{C}_{1-\alpha}(X) \ni \boldsymbol{\vartheta}) \geq 1 - \alpha \quad \forall\, \boldsymbol{\vartheta} \in \Theta.$$

Proof. By Definition 12.1.1,

$$\mathrm{P}_{\boldsymbol{\vartheta}}(\mathcal{C}_{1-\alpha}(X) \ni \boldsymbol{\gamma}) \geq 1-\alpha \quad \forall\, \boldsymbol{\gamma} \in \Gamma_0(\boldsymbol{\vartheta}) \quad \forall\, \boldsymbol{\vartheta} \in \Theta.$$

Because $\gamma(\boldsymbol{\vartheta}) = \boldsymbol{\vartheta}$ is always in $\Gamma_0(\boldsymbol{\vartheta})$ by assumption, the statement follows immediately. □

If $\boldsymbol{\gamma} \in \Theta_0(\boldsymbol{\gamma})$, $\forall\, \boldsymbol{\gamma} \in \Theta$, then also $\boldsymbol{\vartheta} \in \Gamma_0(\boldsymbol{\vartheta})$, $\forall\, \boldsymbol{\vartheta} \in \Theta$, and therefore, the above theorem can be applied to a confidence region corresponding to a test for $H_0 : \boldsymbol{\vartheta} \in \Theta_0(\boldsymbol{\gamma}_0)$ against $H_1 : \boldsymbol{\vartheta} \in \Theta \setminus \Theta_0(\boldsymbol{\gamma}_0)$. Therefore, we can derive a $1-\alpha$ confidence region for $\boldsymbol{\vartheta}$ from any such level α test. This also implies that for the purpose of deriving a $1-\alpha$ confidence region for $\boldsymbol{\vartheta}$ directly from a test, we only have to ensure that the test for the above hypotheses respects the level α at the single point $\boldsymbol{\gamma}_0$, i. e. we can also use a test for a simple null hypothesis against any alternative such that a confidence region of the desired shape results.

If we just wanted to use this method to obtain confidence regions for $\boldsymbol{\vartheta}$, it would therefore suffice to consider confidence regions for $\boldsymbol{\vartheta}$ itself from the beginning by simplifying Definition 12.1.1 using $\gamma(\boldsymbol{\vartheta}) = \boldsymbol{\vartheta}$ and $\Gamma_0(\boldsymbol{\vartheta}) = \{\boldsymbol{\vartheta}\}, \forall\, \boldsymbol{\vartheta} \in \Theta$. However, we will see in the next section that we can take advantage of our more general definition of confidence sets for $\gamma(\boldsymbol{\vartheta})$.

12.4 Sharpened Confidence Regions for $\boldsymbol{\vartheta}$

In the previous section, we obtained confidence regions for $\boldsymbol{\vartheta}$ by simply ignoring the shape of the null parameter region of the underlying test. Of course, we can try to obtain sharper confidence regions by incorporating this information. (If we can exclude certain values of $\boldsymbol{\vartheta}$ due to a priori restrictions on the parameter space, we can sharpen the confidence region by intersecting it with the values that are possible, but this is not a very interesting case.)

In the case of a non-convex $\Theta_0(\boldsymbol{\gamma})$ leading to a non-convex confidence region $\mathcal{C}_{1-\alpha}(X)$ for $\gamma(\boldsymbol{\vartheta})$, it is tempting to sharpen the confidence region for $\boldsymbol{\vartheta}$ by excluding all points that are contained in any $\Theta_0(\boldsymbol{\gamma})$ with $\boldsymbol{\gamma} \notin \mathcal{C}_{1-\alpha}(X)$. However, this usually means that we apply multiple tests to each point, and therefore, this procedure does not guarantee the specified confidence level to hold.

To overcome this multiple testing problem, we can restrict ourselves to a suitable set of $\boldsymbol{\gamma}$'s specified in advance:

Theorem 12.4.1. *Let $\mathcal{C}_{1-\alpha} : \mathcal{X} \to \mathcal{P}(\Gamma)$ be a confidence region for $\gamma(\boldsymbol{\vartheta})$ with confidence level $1-\alpha$ based on the family $(\Gamma_0(\boldsymbol{\vartheta}))_{\boldsymbol{\vartheta}\in\Theta}$. Let $(\Theta_0(\boldsymbol{\gamma}))_{\boldsymbol{\gamma}\in\Gamma}$ be defined as in Theorem 12.1.1. Further, let $\{\boldsymbol{\gamma}_i\}_{i\in I}$, $I \subset \mathbb{R}$, be a subset of the meta-parameter space $\Gamma \subset \mathbb{R}^p$ such that*

$$\Theta_0\left(\boldsymbol{\gamma}_{i_1}\right) \subset \Theta_0\left(\boldsymbol{\gamma}_{i_2}\right) \quad \forall\, i_1, i_2 \in I : i_1 < i_2.$$

As a technical condition, assume that for each subset $\tilde{I} \subset I$, there exists an $i_0 \in I$ such that

$$\bigcap_{i \in \tilde{I}} \Theta_0\left(\boldsymbol{\gamma}_i\right) = \Theta_0\left(\boldsymbol{\gamma}_{i_0}\right).$$

Define

$$\tilde{\mathcal{C}}_{1-\alpha}(X) := \Theta \setminus \bigcup_{i \in I: \boldsymbol{\gamma}_{i'} \notin \mathcal{C}_{1-\alpha}(X)\ \forall\, i' \le i} \Theta_0(\boldsymbol{\gamma}_i).$$

Then

$$\mathrm{P}_{\boldsymbol{\vartheta}}\left(\tilde{\mathcal{C}}_{1-\alpha}(X) \ni \boldsymbol{\vartheta}\right) \ge 1-\alpha \quad \forall\, \boldsymbol{\vartheta} \in \Theta.$$

Proof. Let $(\varphi_i)_{i \in I}$ be the non-randomized tests for $H_{0i} : \boldsymbol{\vartheta} \in \Theta_0(\boldsymbol{\gamma}_i)$ corresponding to $\mathcal{C}_{1-\alpha}$, i. e. $\varphi_i(X) = 1(\mathcal{C}_{1-\alpha}(X) \not\ni \boldsymbol{\gamma}_i)$. By the definition of a confidence set, each φ_i is of level α – see Theorem 12.1.1.

For every $i \in I$, define a test $\tilde{\varphi}_i(X) := \prod_{i' \le i} \varphi_{i'}(X)$ for H_{0i}, which rejects if and only if all $\varphi_{i'}(X)$ with $i' \le i$ reject. We are in a situation as is described in Marcus, Peritz, and Gabriel (1976), and from their argument (and using our technical condition), we can conclude that

$$\mathrm{P}_{\boldsymbol{\vartheta}}\left(\tilde{\varphi}_i(X) = 0\ \forall\, i \in I : \Theta_0(\boldsymbol{\gamma}_i) \ni \boldsymbol{\vartheta}\right) \ge 1-\alpha \quad \forall\, \boldsymbol{\vartheta} \in \Theta.$$

For arbitrary $\boldsymbol{\vartheta} \in \Theta$, it therefore follows that

$$\begin{aligned}
\mathrm{P}_{\boldsymbol{\vartheta}}\left(\tilde{\mathcal{C}}_{1-\alpha}(X) \ni \boldsymbol{\vartheta}\right) &= \mathrm{P}_{\boldsymbol{\vartheta}}\left(\Theta \setminus \bigcup_{i \in I: \boldsymbol{\gamma}_{i'} \notin \mathcal{C}_{1-\alpha}(X)\ \forall\, i' \le i} \Theta_0(\boldsymbol{\gamma}_i) \ni \boldsymbol{\vartheta}\right) \\
&= 1 - \mathrm{P}_{\boldsymbol{\vartheta}}\left(\bigcup_{i \in I: \boldsymbol{\gamma}_{i'} \notin \mathcal{C}_{1-\alpha}(X)\ \forall\, i' \le i} \Theta_0(\boldsymbol{\gamma}_i) \ni \boldsymbol{\vartheta}\right) \\
&= 1 - \mathrm{P}_{\boldsymbol{\vartheta}}\left(\exists\, i \in I : \boldsymbol{\gamma}_{i'} \notin \mathcal{C}_{1-\alpha}(X)\ \forall\, i' \le i, \Theta_0(\boldsymbol{\gamma}_i) \ni \boldsymbol{\vartheta}\right) \\
&= 1 - \mathrm{P}_{\boldsymbol{\vartheta}}\left(\exists\, i \in I : \tilde{\varphi}_i(X) = 1, \Theta_0(\boldsymbol{\gamma}_i) \ni \boldsymbol{\vartheta}\right) \\
&= \mathrm{P}_{\boldsymbol{\vartheta}}\left(\tilde{\varphi}_i(X) = 0\ \forall\, i \in I : \Theta_0(\boldsymbol{\gamma}_i) \ni \boldsymbol{\vartheta}\right) \\
&\ge 1-\alpha.
\end{aligned}$$

□

If I is finite, the technical condition of the above theorem follows from the assumed inclusion of the null parameter regions, and the modified confidence region simplifies to

$$\tilde{\mathcal{C}}_{1-\alpha}(X) = \Theta \setminus \Theta_0\left(\boldsymbol{\gamma}_{i^*}\right) \quad \text{with } i^* = \max\left\{i \in I : \boldsymbol{\gamma}_{i'} \notin \mathcal{C}_{1-\alpha}(X),\ \forall\, i' \le i\right\}$$

(or $\tilde{\mathcal{C}}_{1-\alpha}(X) = \Theta$ if an $i \in I$ with the desired property does not exist).

A possible application of Theorem 12.4.1 is obtained by using (essentially) the straight line $\{(r, \ldots, r)^{\mathrm{T}}\}_{r \in \mathbb{R}} \subset \mathbb{R}^p$ as the meta-parameters to be examined:

Corollary 12.4.2. *Let $\mathcal{C}_{1-\alpha} : \mathcal{X} \to \mathcal{P}(\Gamma)$ be a confidence region for $\gamma(\boldsymbol{\vartheta})$ with confidence level $1-\alpha$ based on the family $(\Gamma_0(\boldsymbol{\vartheta}))_{\boldsymbol{\vartheta}\in\Theta}$. Let $(\Theta_0(\boldsymbol{\gamma}))_{\boldsymbol{\gamma}\in\Gamma}$ be defined as in Theorem 12.1.1. Assume that $\Theta_0(\boldsymbol{\gamma}) = \boldsymbol{\gamma} + \Theta_0(\mathbf{0})$, $\forall\, \boldsymbol{\gamma} \in \Gamma$, that $\Theta_0(\mathbf{0})$ is closed, and that*

$$\Theta_0(\boldsymbol{\gamma}) \subset \Theta_0(\boldsymbol{\gamma} + (\delta, \ldots, \delta)^{\mathrm{T}}) \quad \forall\, \boldsymbol{\gamma} \in \Gamma, \delta > 0.$$

Let ℓ be any real number. Use $\boldsymbol{\gamma}_i = (i, \ldots, i)^{\mathrm{T}} \in \mathbb{R}^p, \forall\, i \in I = [\ell, \infty)$.

For $\tilde{\mathcal{C}}_{1-\alpha}$ as defined in Theorem 12.4.1,

$$\mathrm{P}_{\boldsymbol{\vartheta}}\left(\tilde{\mathcal{C}}_{1-\alpha}(X) \ni \boldsymbol{\vartheta}\right) \geq 1-\alpha \quad \forall\, \boldsymbol{\vartheta} \in \Theta.$$

Proof. The inclusion of the null parameter regions that is needed for the application of Theorem 12.4.1 is obviously fulfilled.

For the technical condition of the theorem, let $\tilde{I} \subset I$. Because $\Theta_0(\boldsymbol{\gamma}) = \boldsymbol{\gamma} + \Theta_0(\mathbf{0})$ and because these regions are closed, we can write

$$\begin{aligned}\bigcap_{i\in\tilde{I}} \Theta_0\left(\boldsymbol{\gamma}_i\right) &= \bigcap_{i\in\tilde{I}} \left(\boldsymbol{\gamma}_i + \Theta_0\left(\mathbf{0}\right)\right) \\ &= \boldsymbol{\gamma}_{i_0} + \Theta_0\left(\mathbf{0}\right) \\ &= \Theta_0(\boldsymbol{\gamma}_{i_0}),\end{aligned}$$

with $i_0 = \inf \tilde{I}$. Since I is closed at the lower end, $i_0 \in I$, and we can apply Theorem 12.4.1. □

An advantage of confidence regions for $\boldsymbol{\vartheta}$ based on Corollary 12.4.2 is that they have the same shape as the alternative parameter region $\Theta \setminus \Theta_0(\mathbf{0})$ of the corresponding test (cf. the example in Figure 12.2). On the other hand, an unpleasant property is that the procedure reduces the multivariate confidence region problem to a univariate problem (a search on the diagonal) and that the set of possible resulting confidence regions is therefore rather restricted.

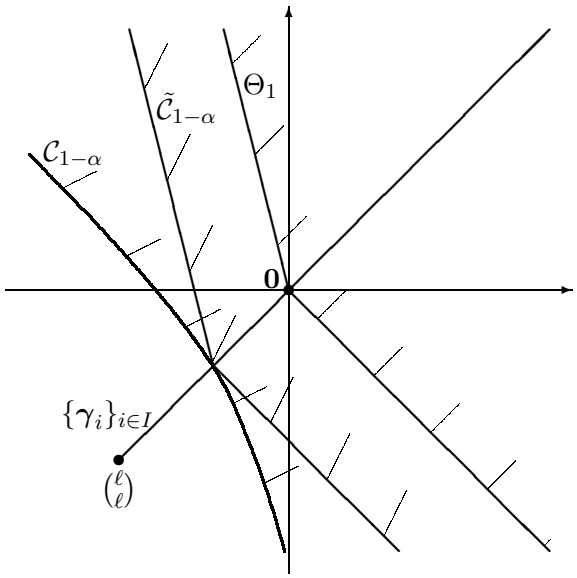

Figure 12.2: Example for the illustration of Corollary 12.4.2: Alternative region $\Theta_1 = \Theta_1(\mathbf{0}) = \mathbb{R}^2 \setminus \Theta_0(\mathbf{0})$, confidence region $\mathcal{C}_{1-\alpha}$ based on the inversion of a test for $H_0 : \boldsymbol{\vartheta} \in \Theta_0(\boldsymbol{\gamma})$ vs. $H_1 : \boldsymbol{\vartheta} \in \mathbb{R}^2 \setminus \Theta_0(\boldsymbol{\gamma})$, and the sharpened confidence region $\tilde{\mathcal{C}}_{1-\alpha}$ obtained from Corollary 12.4.2.

Chapter 13

Example: Pulmonary Function Data

We use an example from the literature to illustrate the application of one-sided tests and the different forms of corresponding confidence regions in the bivariate case.

13.1 Data

The data used is from Merchant et al. (1975), who investigated the effect of exposure to cotton dust on several lung function and blood parameters. Randles (1989), Chakraborty, Chaudhuri, and Oja (1998), and Hettmansperger and McKean (1998, pp. 335ff) use the change in three of these parameters as an example data set to which they apply different multivariate location tests.

We apply bivariate methods to two of these three variables – to the change in forced vital capacity (FVC) and the change in forced expiratory volume (FEV_3), both in liter. These values are given in Table 13.1. In all of the statistical references mentioned, the difference in FVC for subject 11 is given as -0.01, whereas it is -0.10 in the original paper. We also use this modified version of the data for better comparability with these statistical publications.

The conjecture is that the lung function deteriorates under cotton dust exposure, i. e. that the differences in FVC and FEV_3 tend to be negative. It is therefore appropriate to use one-sided location tests and corresponding confidence regions. Different formulations of the exact hypotheses are possible in this example: We could try to show a deterioration in at least one variable, in both variables simultaneously, or in some measure combining both variables.

For consistency with our usual setting, we change the signs of both variables. These data are visualized in Figure 13.1. It is obvious that the two variables are highly correlated; both Pearson's and Spearman's correlation coefficients are 0.90.

Table 13.1: Pulmonary function data: FVC and FEV_3 values used in the example ("statistical version" of the data as in Randles, 1989).

Subject	FVC	FEV_3
1	−0.11	−0.12
2	0.02	0.08
3	−0.02	0.03
4	0.07	0.19
5	−0.16	−0.36
6	−0.42	−0.49
7	−0.32	−0.48
8	−0.35	−0.30
9	−0.10	−0.04
10	0.01	−0.02
11	−0.01	−0.17
12	−0.26	−0.30

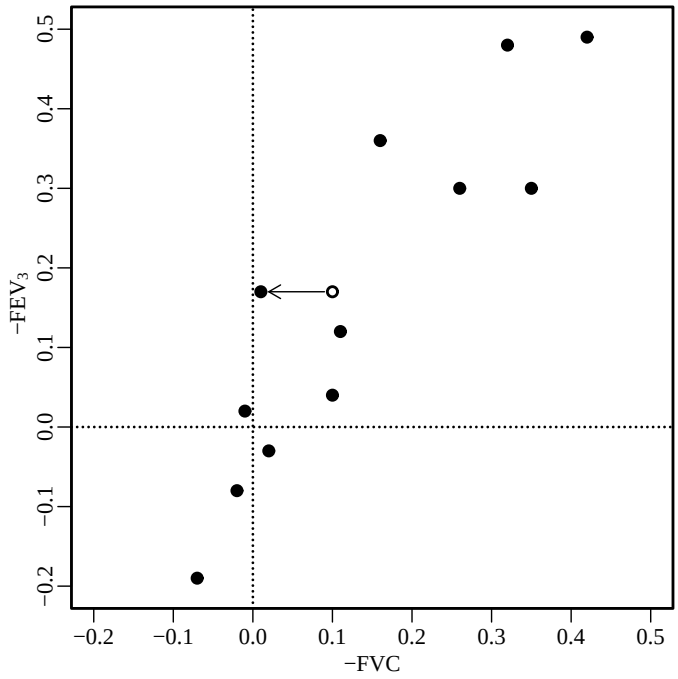

Figure 13.1: Scatter plot of the data in Table 13.1 (dots), with reversed signs. The circle indicates the original data point for subject 11, and the arrow indicates the movement of this point as it entered the statistical literature.

Table 13.2: Test results (p-values) for the data as in Table 13.1 and for the original data from Merchant et al. (1975).

Test	Table 13.1	Original data
sign min test	0.0730	0.0730
Wilcoxon min test	0.0227	0.0227
cone/opposite cone	0.0547	0.0547
$S_{\text{sgn,s}}$	0.0209	0.0252
$S_{\text{Randles,s}}$	0.0260	0.0185
S_{LarLab}	0.0437	0.0437
$S_{\text{LarLab,s}}$	0.0217	0.0217
Brown's angle test	0.0236	0.0219
O'Brien/Wilcoxon	0.0154	0.0092
Perlman (conservative bound)	0.0388	0.0220
t min test	0.0129	0.0129
Glimm/Srivastava/Läuter	0.0216	0.0120
spatial sign test	0.0949	0.0731
Randles's Q_d test	0.0737	0.0344
Blumen's bivariate sign test	0.0727	0.0255
Hotelling's T^2	0.0578	0.0331

13.2 Tests

Table 13.2 shows the results of the tests compared in Chapter 11 when we apply them to these data. The test results are given for the data as in Table 13.1 as well as for the original data set from Merchant et al. (1975).

While the four tests for an unrestricted alternative would not reject the null hypothesis at the level 0.05, most of the tests for restricted alternatives would. Notice also that the effect of changing one point is considerable for many tests: Three of the four tests for an unrestricted alternative would have rejected at the level 0.05 for the original data, and even some of the (allegedly more robust) nonparametric methods are remarkably affected. The min tests are not affected – they benefit from the fact that the change only happened in the "less significant" variable. The cone/opposite cone test does not change its p-value either because the moved point stays within the first quadrant, and S_{LarLab} and $S_{\text{LarLab,s}}$, too, lead to the same p-value for both versions of the data.

13.3 Confidence Regions

We can numerically invert these tests at the nominal level α in order to obtain nominal $1-\alpha$ confidence regions. We do this by testing $H_0 : \boldsymbol{\vartheta} = \boldsymbol{\gamma}$ for each

point $\boldsymbol{\gamma}$ on a grid. Because each of the tests used respects the level α for this simple null hypothesis, we can interpret the confidence regions thus obtained as $1 - \alpha$ confidence regions for $\boldsymbol{\vartheta}$ – see Section 12.3. In Figure 13.2, the borders of the confidence regions obtained using a grid of $\{-1, -0.99, -0.98, \ldots, 1\}^2$ and $\alpha = 0.05$ are shown; each class of tests according to the list on page 67 is shown in a separate panel of the figure.

The confidence region for the meta-parameter $\boldsymbol{\gamma}$ is the region inside the curve for the first class of tests (unrestricted alternative tests, part (a) of Figure 11.3). Whereas Hotelling's T^2 produces an elliptical confidence region, the confidence regions of the three nonparametric tests are all star-shaped and can hardly be distinguished in the plot.

The confidence region corresponding to each of the restricted alternative tests is the region above and to the right of the curve, i. e. it contains the upper right corner of the plot. It is instructive to compare these confidence regions to the curves of power 0.05 in Figure 11.3 (b). (Our data do not appear to be too far away from a normal distribution, and they are strongly positively correlated.) The shape of the confidence region is always quite similar to $-\tilde{\Theta}_0$ when we define $\tilde{\Theta}_0$ as the region where the power is below 0.05. Deviations occur mainly in the central portion of the border. When we interpret $\tilde{\Theta}_0$ as the "idealized" null parameter region for some test under the given distribution, we see that a result similar to Theorem 12.2.5 seems to hold under far more general conditions: E. g., the conservative tests for a three-quadrant null hypothesis vs. a one-quadrant alternative (Figure 13.2 (f)) produce confidence regions that are translations of the complement of the negative quadrant, i. e. translations of $-\Theta_0$. This follows from Corollary 12.2.2. But also for the liberal tests for the same hypotheses (Figure 13.2 (e)), the shape of the confidence region only slightly deviates from a suitable translation of $-\Theta_0$, and the border of the confidence region seems to converge to this shape when moving away from the origin. The shape of the confidence region is again very similar to that of $-\tilde{\Theta}_0$. With $-\tilde{\Theta}_0$, analogous observations can also be made for the other classes of tests. E. g., for Perlman's test, the confidence region does not resemble a translated version of $-\Theta_0$, which would be the positive quadrant. But this test does not maintain the specified level anyhow for this composite null hypothesis in the positively correlated case, and again, the confidence region has a shape very similar to that of $-\tilde{\Theta}_0$, as can both be seen from Figure 11.3 (b).

From Section 12.3, we know that we only have to guarantee the level of a test to be inverted at the simple null hypothesis $\boldsymbol{\vartheta} = \mathbf{0}$. It would therefore be plausible that the best (i. e. smallest in some sense) confidence regions are obtained by inverting tests that (approximately) reach their nominal level α at the simple null hypothesis. However, we can see that the confidence regions in Figure 13.2 (e) are at least not uniformly better than e. g. that of the Wilcoxon min test, even though they are based on tests with a level near α for the simple

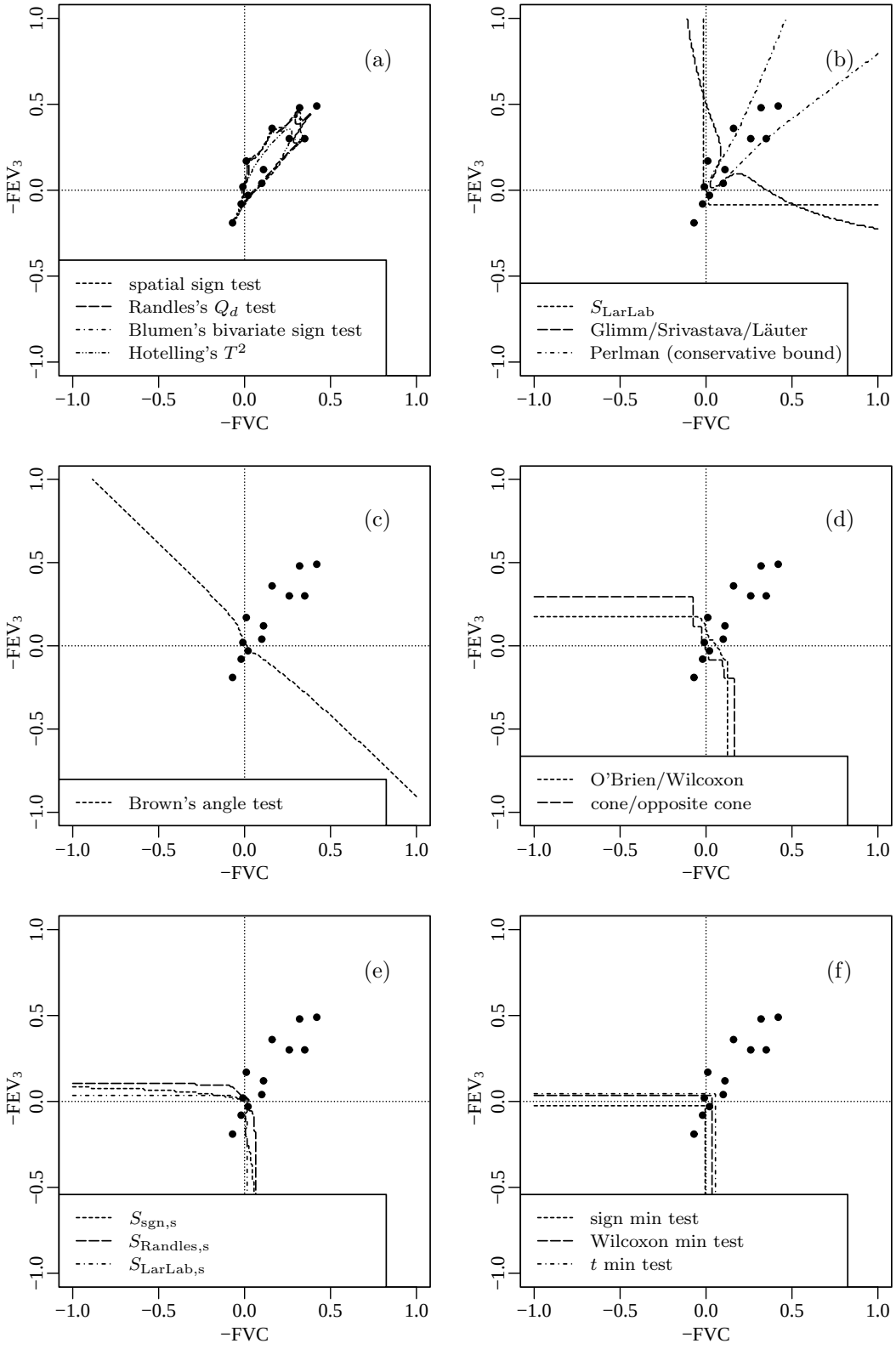

Figure 13.2: 95% confidence regions according to Section 12.3.

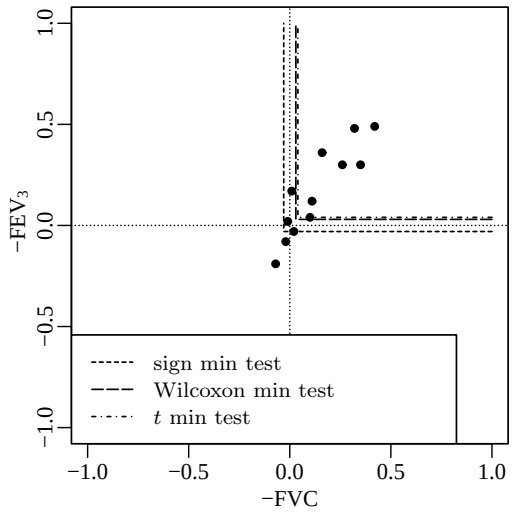

Figure 13.3: Sharpened 95% confidence regions according to Corollary 12.4.2.

null hypothesis. Especially if our main interest is in obtaining a confidence region that does not include any point in the third quadrant, the confidence region based on the Wilcoxon min test could be a good choice, even though this test does not attain the nominal level at the simple null hypothesis.

Finally, we look at the sharpened confidence regions obtained from the application of Corollary 12.4.2 to the confidence regions obtained from inverting the min tests. These are given in Figure 13.3. Here, the shape of the confidence region corresponds to that of the alternative parameter region $\mathbb{R}^2_+$, and at least the sharpened confidence regions obtained from the Wilcoxon and t min tests are uniformly better than the confidence region obtained by inverting the test based on S_{LarLab} (see Figure 13.2 (b)), which has a similar shape. In this example, the sharpened confidence regions benefit from the fact that the data points are situated rather close to the diagonal.

Chapter 14

Discussion and Outlook

A large part of the work for this thesis consisted in studying articles from the areas of multivariate statistics, nonparametric methods, multiple testing, and other related subjects, in summarizing and combining the relevant results, as well as in the implementation of methods in the R programming language. As independent and novel contributions, the following aspects are worth mentioning, and most of them suggest possibilities for further work:

- Several novel one-sample test procedures were proposed for multivariate restricted alternative problems, with a focus on convex sector and cone alternatives.

- The graphical method proposed in Section 10.3 seems to be useful for evaluating tests for a specific combination of composite null and alternative hypotheses. Further work could involve a more detailed study of possible applications in higher dimensions.

  An algorithm for the estimation of the curve of constant power (in the bivariate case) is given in Appendix C. While this algorithm worked reasonably well for the simulations done in Chapter 11, it would be interesting to investigate its theoretical properties and possibly to develop a faster and/or more accurate algorithm. Especially for the investigation of bootstrap tests (as those in Minhajuddin, Frawley, Schucany, and Woodward, 2006), a faster algorithm would be very useful.

- The simulations from Chapter 11 were used to obtain a classification of tests. Although simulations with other distributions, parameters, and sample sizes should be done to verify this classification, it does not seem to vary too much as long as the assumptions for the tests used are fulfilled. Different alternative parameter regions could also be investigated for the tests that allow for a general sector alternative.

- In Chapter 12, a flexible notation for confidence regions corresponding to restricted alternative tests was given. This notation allowed for statements about the shape of confidence regions for multivariate location parameters and also for the construction of refined confidence regions in Section 12.4.

  As there does not seem to be any similar work in the literature, many extensions can be thought of: Sharper results about the shape of a confidence region corresponding to a test in dimension $p > 2$ can be expected, and Theorem 12.2.5 (for the bivariate case) is likely to hold under less stringent assumptions, as can be guessed from the example in Chapter 13. Finally, there will certainly be alternatives to the method in Section 12.4 for obtaining confidence regions for $\boldsymbol{\vartheta}$ itself that are of a desirable shape.

Since our focus was on the one-sample case, an obvious area for further research consists in the development of similar methods for two- and multi-sample cases.

Acknowledgements

My thanks go to Prof. Dr. Jürg Hüsler for proposing me to work on this subject, for helpful suggestions, and for providing me the opportunity to choose my focuses myself. I am also grateful to Prof. Dr. Thomas P. Hettmansperger for the interesting discussion about several aspects of my work on the occasion of his visit at the University of Bern. Further, I thank Prof. Dr. Lutz Dümbgen and Prof. Dr. Ilya Molchanov, as well as my colleagues Marco Collenberg, Michael Mayer, Dr. Samuel Müller, and Dr. Kaspar Rufibach for helpful scientific hints.

I also thank all the members of the IMSV for the pleasant atmosphere – special thanks go to my colleagues who participated in semi-scientific and non-scientific discussions, common lunches, and playing cards. Last but not least, my thanks go to Corinne and to my family, especially for encouraging me in times when there was hardly any progress in my work.

This work was supported by the Swiss National Science Foundation.

Bern, November 2006
Michael Vock

Appendix A

Symmetry Centers and Expectations

A.1 Interpretation of Symmetry Centers Using Expectations

Theorem A.1.1. *Let* $\boldsymbol{X}$ *have a diagonally symmetric distribution with respect to* $\boldsymbol{\vartheta}$*, and let* $\mathrm{E}(\boldsymbol{X})$ *exist.*

Then

$$\mathrm{E}(\boldsymbol{X}) = \boldsymbol{\vartheta}.$$

Proof. Diagonal symmetry with respect to $\boldsymbol{\vartheta}$ means that $\boldsymbol{X} - \boldsymbol{\vartheta} \overset{\mathrm{d}}{=} \boldsymbol{\vartheta} - \boldsymbol{X}$. Therefore,

$$\mathrm{E}(\boldsymbol{X}) - \boldsymbol{\vartheta} = \mathrm{E}(\boldsymbol{X} - \boldsymbol{\vartheta}) = \mathrm{E}(\boldsymbol{\vartheta} - \boldsymbol{X}) = \boldsymbol{\vartheta} - \mathrm{E}(\boldsymbol{X}),$$

such that $2\,\mathrm{E}(\boldsymbol{X}) = 2\boldsymbol{\vartheta}$. □

Corollary A.1.2. *Let* $\boldsymbol{X}$ *have a directionally symmetric distribution with respect to* $\boldsymbol{\vartheta}$*.*

Then

$$\mathrm{E}(\operatorname{sgn} \boldsymbol{X} - \boldsymbol{\vartheta}) = \mathbf{0}.$$

Proof. Directional symmetry of $\boldsymbol{X}$ with respect to $\boldsymbol{\vartheta}$ implies diagonal symmetry of $\operatorname{sgn} \boldsymbol{X} - \boldsymbol{\vartheta}$ with respect to $\mathbf{0}$. As the expected spatial sign always exists, we can apply Theorem A.1.1 and immediately obtain the result. □

By these results, the simple null hypothesis $H_0 : \boldsymbol{\vartheta} = \mathbf{0}$ implies $\mathrm{E}(\boldsymbol{X}) = \mathbf{0}$ (under diagonal symmetry and existence of the expectation) and $\mathrm{E}(\operatorname{sgn} \boldsymbol{X}) = \mathbf{0}$ (under directional symmetry). Therefore, a rejection of one of these more specialized hypotheses can also be interpreted as a rejection of H_0.

A.2 Direction of the Expected Spatial Sign

The hypotheses for the one-sided versions of the spatial sign test (Chapter 7) and of the sign test by Randles (Chapter 8) are formulated in terms of expected spatial signs. While the results from the previous section are helpful in the interpretation of the null hypotheses of these tests, we would also like to have some knowledge about the relationship between the expected spatial sign and the expectation of the original observations under $\boldsymbol{\vartheta} \neq \mathbf{0}$ for the comparison with other tests.

Theorem A.2.1. *Let $\boldsymbol{X}$ be a random vector in $\mathbb{R}^p$, and let its distribution F be directionally symmetric with respect to $\boldsymbol{\vartheta} = \mathrm{E}(\boldsymbol{X}) \neq \mathbf{0}$.*

If $\mathrm{E}(\operatorname{sgn} \boldsymbol{X}) \neq \mathbf{0}$, then

$$\mathrm{E}(\operatorname{sgn} \boldsymbol{X})^{\mathrm{T}} \mathrm{E}(\boldsymbol{X}) > 0.$$

Proof. Without loss of generality, assume that $\boldsymbol{\vartheta} = (\vartheta_1, 0, \ldots, 0)^{\mathrm{T}}$ with $\vartheta_1 > 0$ (by a rotation of the coordinate system). Consider the conditional expectation of $\operatorname{sgn} \boldsymbol{X}$ given that $\boldsymbol{X}$ lies on some line through $\boldsymbol{\vartheta}$ specified by $\boldsymbol{v} \in \mathbb{R}_+ \times \mathbb{R}^{p-1}$, with $\|\boldsymbol{v}\| = 1$ and $\boldsymbol{v} \neq \operatorname{sgn} \boldsymbol{\vartheta}$. (We are only interested in lines through $\boldsymbol{\vartheta}$ that $\boldsymbol{X}$ can lie on, and for these lines, the conditional expectation exists.) The most important step of the proof is to show that the first component of this conditional expectation is greater than 0. By the directional symmetry,

$$\begin{aligned} &\mathrm{E}(\operatorname{sgn} \boldsymbol{X} | \boldsymbol{X} \in \{\boldsymbol{\vartheta} + c\boldsymbol{v} : c \in \mathbb{R} \setminus \{0\}\}) \\ &\quad = \frac{1}{2} \left[\mathrm{E}(\operatorname{sgn} \boldsymbol{X} | \boldsymbol{X} \in \{\boldsymbol{\vartheta} + c\boldsymbol{v} : c > 0\}) + \mathrm{E}(\operatorname{sgn} \boldsymbol{X} | \boldsymbol{X} \in \{\boldsymbol{\vartheta} + c\boldsymbol{v} : c < 0\}) \right]. \end{aligned} \tag{$*$}$$

Let $\varphi = \arccos \frac{\boldsymbol{v}^{\mathrm{T}} \boldsymbol{\vartheta}}{\|\boldsymbol{\vartheta}\|} = \arccos v_1 \in (0, \frac{\pi}{2})$ be the angle between $\boldsymbol{\vartheta}$ and $\boldsymbol{v}$. Further, let

$$\alpha = \arccos \frac{\mathrm{E}(\operatorname{sgn} \boldsymbol{X} | \boldsymbol{X} \in R_{\boldsymbol{v}})^{\mathrm{T}} \boldsymbol{\vartheta}}{\|\mathrm{E}(\operatorname{sgn} \boldsymbol{X} | \boldsymbol{X} \in R_{\boldsymbol{v}})\| \, \|\boldsymbol{\vartheta}\|}$$

be the angle between $\boldsymbol{\vartheta}$ and the expected spatial sign given $\boldsymbol{X} \in R_{\boldsymbol{v}} = \{\boldsymbol{\vartheta} + c\boldsymbol{v} : c > 0\}$. Necessarily, $\alpha \in (0, \varphi)$. Moreover, $\mathrm{E}(\operatorname{sgn} \boldsymbol{X} | \boldsymbol{X} \in R_{\boldsymbol{v}})$ is within the circular segment shown in Figure A.1. From this figure, we can see that for a given α, the norm of $\mathrm{E}(\operatorname{sgn} \boldsymbol{X} | \boldsymbol{X} \in R_{\boldsymbol{v}})$ tends towards its infimum if $\boldsymbol{X} = \boldsymbol{\vartheta} + \frac{1}{c}\boldsymbol{v}$ with some probability $\frac{q}{q+1}$ and $\boldsymbol{X} = \boldsymbol{\vartheta} + c\boldsymbol{v}$ with probability $\frac{1}{q+1}$, where $c \to \infty$. In the limiting case, $\operatorname{sgn}(\boldsymbol{\vartheta} + \frac{1}{c}\boldsymbol{v}) \to \operatorname{sgn} \boldsymbol{\vartheta} = (1, 0, \ldots, 0)^{\mathrm{T}}$, and $\operatorname{sgn}(\boldsymbol{\vartheta} + c\boldsymbol{v}) \to \boldsymbol{v} = (\cos \varphi, v_2, \ldots, v_p)^{\mathrm{T}}$, with

$$\sqrt{v_2^2 + \ldots + v_p^2} = \sqrt{1 - (\cos \varphi)^2} = \sin \varphi.$$

q has to satisfy

$$\tan \alpha = \frac{\frac{q}{q+1} \cdot 0 + \frac{1}{q+1} \sin \varphi}{\frac{q}{q+1} \cdot 1 + \frac{1}{q+1} \cos \varphi} = \frac{\sin \varphi}{q + \cos \varphi}.$$

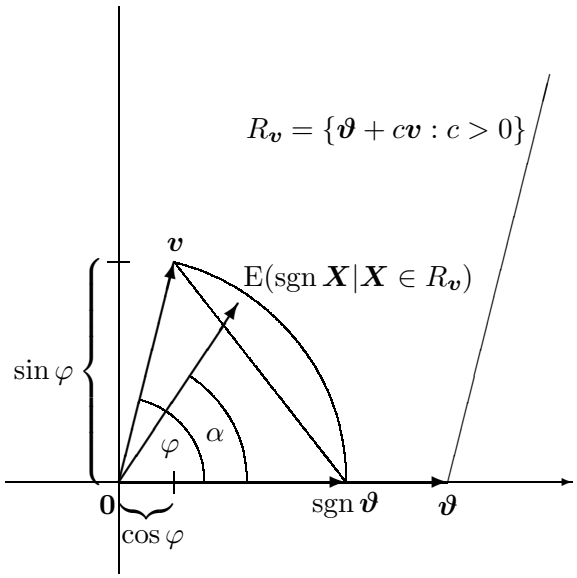

Figure A.1: Notation for the proof of Theorem A.2.1. The figure shows the plane spanned by $\boldsymbol{\vartheta}$ and $\boldsymbol{v}$.

Therefore, $q = \frac{\sin\varphi}{\tan\alpha} - \cos\varphi$. This yields the following lower bound for the first component of $\mathrm{E}(\operatorname{sgn} \boldsymbol{X} | \boldsymbol{X} \in R_{\boldsymbol{v}})$:

$$\begin{aligned}
\mathrm{E}(\operatorname{sgn} \boldsymbol{X} | \boldsymbol{X} \in R_{\boldsymbol{v}})_1 &> \frac{q}{q+1} \cdot 1 + \frac{1}{q+1} \cos\varphi = \frac{1}{q+1}(q + \cos\varphi) \\
&= \frac{1}{\frac{\sin\varphi}{\tan\alpha} - \cos\varphi + 1} \cdot \frac{\sin\varphi}{\tan\alpha} = \frac{\sin\varphi}{\sin\varphi + \tan\alpha(1 - \cos\varphi)} \\
&> \frac{\sin\varphi}{\sin\varphi + \tan\varphi(1 - \cos\varphi)} = \frac{1}{1 + \frac{1-\cos\varphi}{\cos\varphi}} = \cos\varphi
\end{aligned}$$

A simple lower bound for the first component of the second term on the right-hand side of $(*)$ is

$$\mathrm{E}(\operatorname{sgn} \boldsymbol{X} | \boldsymbol{X} \in \{\boldsymbol{\vartheta} + c\boldsymbol{v} : c < 0\})_1 > \cos(\varphi + \pi) = -\cos\varphi,$$

and therefore $(*)$ yields

$$\mathrm{E}(\operatorname{sgn} \boldsymbol{X} | \boldsymbol{X} \in \{\boldsymbol{\vartheta} + c\boldsymbol{v} : c \in \mathbb{R} \setminus \{0\}\})_1 > \frac{1}{2}\left[\cos\varphi - \cos\varphi\right] = 0.$$

We can easily deal with the remaining special cases:

$$\begin{aligned}
&\mathrm{E}(\operatorname{sgn} \boldsymbol{X} | \boldsymbol{X} \in \{c\boldsymbol{\vartheta} : c \in \mathbb{R}\})_1 \\
&\quad = \mathrm{P}(X_1 > 0 | \boldsymbol{X} \in \{c\boldsymbol{\vartheta} : c \in \mathbb{R}\}) - \mathrm{P}(X_1 < 0 | \boldsymbol{X} \in \{c\boldsymbol{\vartheta} : c \in \mathbb{R}\}) \geq 0
\end{aligned}$$

because of the choice of the coordinate system, with equality holding only if $\mathrm{E}(\operatorname{sgn}\boldsymbol{X}|\boldsymbol{X} \in \{c\boldsymbol{\vartheta} : c \in \mathbb{R}\}) = \mathbf{0}$. Further, for $\boldsymbol{v} \in \{0\} \times \mathbb{R}^{p-1}$, $\mathrm{E}(\operatorname{sgn}\boldsymbol{X}|\boldsymbol{X} \in \{\boldsymbol{\vartheta} + c\boldsymbol{v} : c \in \mathbb{R} \setminus \{0\}\})$ trivially has a positive first component.

Integration over the condition yields $\mathrm{E}(\operatorname{sgn}\boldsymbol{X})_1 > 0$ (equality could only be obtained with $\mathrm{P}(\boldsymbol{X} \in \{c\boldsymbol{\vartheta} : c \in \mathbb{R}\}) = 1$ and $\mathrm{E}(\operatorname{sgn}\boldsymbol{X}|\boldsymbol{X} \in \{c\boldsymbol{\vartheta} : c \in \mathbb{R}\}) = \mathbf{0}$, but then $\mathrm{E}(\operatorname{sgn}\boldsymbol{X}) = \mathbf{0}$, which is excluded by assumption). Therefore

$$\mathrm{E}(\operatorname{sgn}\boldsymbol{X})^{\mathrm{T}}\,\mathrm{E}(\boldsymbol{X}) = \mathrm{E}(\operatorname{sgn}\boldsymbol{X})_1\vartheta_1 > 0. \qquad \square$$

According to Theorem A.2.1, the directions of $\mathrm{E}(\boldsymbol{X})$ and $\mathrm{E}(\operatorname{sgn}\boldsymbol{X})$ deviate by an angle of less than $\frac{\pi}{2}$ for directionally symmetric distributions. The following example shows that no sharper bound can be given for this general situation.

Example A.2.1. Let $\boldsymbol{X}$ be a bivariate random vector taking the values $(3,\varepsilon)^{\mathrm{T}}$ and $(-1,\varepsilon)^{\mathrm{T}}$, each with probability $\frac{1}{2}$, for some $\varepsilon > 0$. $\boldsymbol{X}$ is directionally symmetric with respect to $(0,\varepsilon)^{\mathrm{T}}$, and obviously, $\mathrm{E}(\boldsymbol{X}) = (1,\varepsilon)^{\mathrm{T}} \to (1,0)^{\mathrm{T}}$ for $\varepsilon \to 0$.

The possible values of $\operatorname{sgn}\boldsymbol{X}$, again each occurring with probability $\frac{1}{2}$, are $(3/\sqrt{9+\varepsilon^2}, \varepsilon/\sqrt{9+\varepsilon^2})^{\mathrm{T}}$ and $(-1/\sqrt{1+\varepsilon^2}, \varepsilon/\sqrt{1+\varepsilon^2})^{\mathrm{T}}$. Therefore, $\mathrm{E}(\operatorname{sgn}\boldsymbol{X}) = c(3\sqrt{1+\varepsilon^2} - \sqrt{9+\varepsilon^2}, \varepsilon\sqrt{1+\varepsilon^2} + \varepsilon\sqrt{9+\varepsilon^2})^{\mathrm{T}}$, with $c = 1/(2\sqrt{1+\varepsilon^2}\sqrt{9+\varepsilon^2})$.

Let $\beta \in [0, \pi/2)$ be the angle between the first coordinate axis and $\mathrm{E}(\operatorname{sgn}\boldsymbol{X})$. By the Maclaurin expansion $\sqrt{1+x} = 1 + \frac{x}{2} + \mathrm{O}(x^2)$ $(x \to 0)$,

$$\begin{aligned}\tan\beta &= \frac{\varepsilon\sqrt{1+\varepsilon^2} + \varepsilon\sqrt{9+\varepsilon^2}}{3\sqrt{1+\varepsilon^2} - \sqrt{9+\varepsilon^2}} = \frac{\varepsilon\left(1 + \frac{\varepsilon^2}{2} + 3 + \frac{3\varepsilon^2}{18} + \mathrm{O}(\varepsilon^4)\right)}{3 + \frac{3\varepsilon^2}{2} - 3 - \frac{3\varepsilon^2}{18} + \mathrm{O}(\varepsilon^4)} \\ &= \frac{\varepsilon\,(4 + \mathrm{O}(\varepsilon^2))}{\varepsilon^2\left(\frac{4}{3} + \mathrm{O}(\varepsilon^2)\right)} \to \infty \quad (\varepsilon \to 0),\end{aligned}$$

and $\beta \to \frac{\pi}{2}$. Thus, the directions of $\mathrm{E}(\boldsymbol{X})$ and $\mathrm{E}(\operatorname{sgn}\boldsymbol{X})$ deviate by an angle approaching $\frac{\pi}{2}$ as $\varepsilon \to 0$. $\triangle$

We can also find (restrictive) symmetry conditions that guarantee the directions of $\mathrm{E}(\boldsymbol{X})$ and $\mathrm{E}(\operatorname{sgn}\boldsymbol{X})$ to coincide:

Theorem A.2.2. *Let $\boldsymbol{X}$ be a random vector in $\mathbb{R}^p$, and let its distribution F be diagonally symmetric with respect to $\boldsymbol{\vartheta} = \mathrm{E}(\boldsymbol{X}) \neq \mathbf{0}$. Further, let F be symmetric with respect to the hyperplane through the point $\boldsymbol{\vartheta}$ that is orthogonal to the vector $\boldsymbol{\vartheta}$, in the sense that*

$$dF(\boldsymbol{x}) = dF\left(\boldsymbol{x} - 2\left(\frac{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\vartheta}}{\|\boldsymbol{\vartheta}\|} \cdot \frac{\boldsymbol{\vartheta}}{\|\boldsymbol{\vartheta}\|} - \boldsymbol{\vartheta}\right)\right), \forall\, \boldsymbol{x} \in \mathbb{R}^p.$$

Then

$$\mathrm{E}(\operatorname{sgn}\boldsymbol{X}) = c\,\mathrm{E}(\boldsymbol{X}) = c\boldsymbol{\vartheta}$$

for some $c > 0$.

Proof. Without loss of generality, as in the proof of Theorem A.2.1, assume $\boldsymbol{\vartheta} = (\vartheta_1, 0, \ldots, 0)^{\mathrm{T}}$ with $\vartheta_1 > 0$. The second symmetry condition therefore simplifies to

$$dF(\boldsymbol{x}) = dF\left((2\vartheta_1 - x_1, x_2, \ldots, x_p)^{\mathrm{T}}\right), \forall\, \boldsymbol{x} \in \mathbb{R}^p.$$

Combining this with the diagonal symmetry with respect to $\boldsymbol{\vartheta}$, we obtain

$$dF(\boldsymbol{x}) = dF\left((x_1, -x_2, \ldots, -x_p)^{\mathrm{T}}\right), \forall\, \boldsymbol{x} \in \mathbb{R}^p,$$

i. e. symmetry with respect to the first coordinate axis, or, in the general case, the straight line given by $\boldsymbol{\vartheta}$. In other words, conditional on every hyperplane that is orthogonal on $\boldsymbol{\vartheta}$, the distribution is diagonally symmetric with respect to the point of intersection of the hyperplane and the line given by $\boldsymbol{\vartheta}$. Therefore, $\mathrm{E}(\operatorname{sgn} \boldsymbol{X} | X_1 = t) = c(t)\boldsymbol{\vartheta}$ for some real-valued function $c(t)$, with $\operatorname{sgn} c(t) = \operatorname{sgn} t$.

Due to the diagonal symmetry assumption, the contribution of some point $\boldsymbol{x}$ with $x_1 = t > \vartheta_1$ to $\mathrm{E}(\operatorname{sgn} \boldsymbol{X} | X_1 = t)$ is equally important as that of $2\boldsymbol{\vartheta} - \boldsymbol{x}$ to $\mathrm{E}(\operatorname{sgn} \boldsymbol{X} | X_1 = 2\vartheta_1 - t)$. Given that only the first component is nonzero in these conditional expectations, we can focus on the first component of each contribution. Under $x_1 = t > \vartheta_1$, by elementary geometrical considerations,

$$(\operatorname{sgn} \boldsymbol{x})_1 = \frac{t}{\sqrt{t^2 + \|\boldsymbol{x}\|^2 - t^2}} > \frac{|2\vartheta_1 - t|}{\sqrt{(2\vartheta_1 - t)^2 + \|\boldsymbol{x}\|^2 - t^2}} = \left|(\operatorname{sgn}(2\boldsymbol{\vartheta} - \boldsymbol{x}))_1\right|,$$

the inequality being due to the monotonicity of $t \mapsto t/\sqrt{t^2 + a^2}$ and the fact that $|2\vartheta_1 - t| < t$. Therefore, still for $t > \vartheta_1$, $c(t) > |c(2\vartheta_1 - t)|$, such that

$$c(t) + c(2\vartheta_1 - t) > 0. \qquad (*)$$

Now

$$\begin{aligned}
&\mathrm{E}(\operatorname{sgn} \boldsymbol{X}) \\
&\quad = \int_0^\infty \mathrm{E}(\operatorname{sgn} \boldsymbol{X} | X_1 = \vartheta_1 + s)\, d\,\mathrm{P}(X_1 = \vartheta_1 + s) \\
&\qquad + \int_0^\infty \mathrm{E}(\operatorname{sgn} \boldsymbol{X} | X_1 = \vartheta_1 - s)\, d\,\mathrm{P}(X_1 = \vartheta_1 - s) \\
&\quad = \int_0^\infty \frac{1}{2}[\mathrm{E}(\operatorname{sgn} \boldsymbol{X} | X_1 = \vartheta_1 + s) + \mathrm{E}(\operatorname{sgn} \boldsymbol{X} | X_1 = \vartheta_1 - s)]\, d\,\mathrm{P}(|X_1 - \vartheta_1| = s) \\
&\quad = \int_0^\infty \frac{1}{2}[c(\vartheta_1 + s) + c(\vartheta_1 - s)]\boldsymbol{\vartheta}\, d\,\mathrm{P}(|X_1 - \vartheta_1| = s),
\end{aligned}$$

and according to $(*)$, the expression in brackets is positive, which completes the proof. □

Appendix B

Underlying Results

B.1 Asymptotic Distribution of Random Vectors

Theorem B.1.1. *Let $(\boldsymbol{X}_n)_{n\in\mathbb{N}}$ be a sequence of random vectors in $\mathbb{R}^p$, $(\boldsymbol{Y}_n)_{n\in\mathbb{N}}$ a sequence of random vectors in $\mathbb{R}^q$, $\boldsymbol{X}_n \xrightarrow{\mathrm{d}} \boldsymbol{X}$, and $\boldsymbol{Y}_n \xrightarrow{\mathrm{d}} \boldsymbol{Y}$. Further, let one of the following conditions hold:*

(a) *$\boldsymbol{Y}$ is a. s. a constant, or*

(b) *$\boldsymbol{X}_n$ and $\boldsymbol{Y}_n$ are stochastically independent for each $n \in \mathbb{N}$.*

Then $(\boldsymbol{X}_n, \boldsymbol{Y}_n) \xrightarrow{\mathrm{d}} (\boldsymbol{X}, \boldsymbol{Y})$, with $F_{(\boldsymbol{X},\boldsymbol{Y})}(\boldsymbol{x}, \boldsymbol{y}) = F_{\boldsymbol{X}}(\boldsymbol{x})F_{\boldsymbol{Y}}(\boldsymbol{y})$.

Proof. See Witting and Nölle (1970), Hilfssatz 2.5. □

Theorem B.1.2. *Let $(\boldsymbol{X}_n)_{n\in\mathbb{N}}$ be a sequence of random vectors in $\mathbb{R}^p$ and $g : \mathbb{R}^p \to \mathbb{R}^q$ be continuous at every point in a set C such that $\mathrm{P}(\boldsymbol{X} \in C) = 1$.*

(a) *Let $\boldsymbol{X}_n \xrightarrow{\mathrm{d}} \boldsymbol{X}$. Then $g(\boldsymbol{X}_n) \xrightarrow{\mathrm{d}} g(\boldsymbol{X})$.*

(b) *Let $\boldsymbol{X}_n \xrightarrow{\mathrm{p}} \boldsymbol{X}$. Then $g(\boldsymbol{X}_n) \xrightarrow{\mathrm{p}} g(\boldsymbol{X})$.*

Proof. See van der Waart (1998), Theorem 2.3 (i), (ii). □

Theorem B.1.3. *Let $(\boldsymbol{X}_n)_{n\in\mathbb{N}}, (\boldsymbol{Y}_n)_{n\in\mathbb{N}}$ be sequences of random vectors in $\mathbb{R}^p$, $\boldsymbol{X}_n \xrightarrow{\mathrm{d}} \boldsymbol{X}$, and $\boldsymbol{X}_n - \boldsymbol{Y}_n \xrightarrow{\mathrm{p}} \boldsymbol{0}$. Then $\boldsymbol{Y}_n \xrightarrow{\mathrm{d}} \boldsymbol{X}$.*

Proof. See van der Waart (1998), Theorem 2.7 (iv). □

Theorem B.1.4. *Let $(\boldsymbol{X}_n)_{n\in\mathbb{N}}$ be a sequence of random vectors in $\mathbb{R}^p$, $\boldsymbol{X}_n \xrightarrow{\mathrm{d}} \mathrm{N}_p(\boldsymbol{0}, \Sigma)$, Σ positive definite, $\Sigma_n^{-1} \xrightarrow{\mathrm{p}} \Sigma^{-1}$. Then:*

(a) $\Sigma^{-\frac{1}{2}} \boldsymbol{X}_n \xrightarrow{\mathrm{d}} \mathrm{N}_p(\mathbf{0}, \mathrm{I}_p)$

(b) $\boldsymbol{X}_n^{\mathrm{T}} \Sigma^{-1} \boldsymbol{X}_n \xrightarrow{\mathrm{d}} \chi^2(p)$

(c) $\Sigma_n^{-\frac{1}{2}} \boldsymbol{X}_n \xrightarrow{\mathrm{d}} \mathrm{N}_p(\mathbf{0}, \mathrm{I}_p)$

(d) $\boldsymbol{X}_n^{\mathrm{T}} \Sigma_n^{-1} \boldsymbol{X}_n \xrightarrow{\mathrm{d}} \chi^2(p)$

Proof. See Fahrmeir and Hamerle (1984), Satz 3.11. □

B.2 Inequalities for Binomial Probabilities

Theorem B.2.1. *Let B_n denote a random variable from a binomial distribution with parameters n and 0.5, with $n \geq 1$.*

For $k \in \{0, \ldots, n\}$,

$$\mathrm{P}(B_{n+1} \geq k+1) < \mathrm{P}(B_n \geq k) \leq \mathrm{P}(B_{n+1} \geq k).$$

Proof. Using the well-known relationship

$$\binom{n+1}{k+1} = \binom{n}{k} + \binom{n}{k+1},$$

we obtain

$$\begin{aligned}
\mathrm{P}(B_n \geq k) &= \frac{1}{2^n} \sum_{\ell=k}^{n} \binom{n}{\ell} \\
&= \frac{1}{2^{n+1}} \left[\binom{n}{k} + \sum_{\ell=k}^{n-1} \binom{n}{\ell} + \sum_{\ell=k+1}^{n} \binom{n}{\ell} + \binom{n}{n} \right] \\
&= \frac{1}{2^{n+1}} \left[\binom{n}{k} + \sum_{\ell=k+1}^{n} \left(\binom{n}{\ell-1} + \binom{n}{\ell} \right) + 1 \right] \\
&= \frac{1}{2^{n+1}} \left[\binom{n}{k} + \sum_{\ell=k+1}^{n} \binom{n+1}{\ell} + \binom{n+1}{n+1} \right] \\
&= \frac{1}{2^{n+1}} \left[\binom{n}{k} + \sum_{\ell=k+1}^{n+1} \binom{n+1}{\ell} \right].
\end{aligned}$$

$\binom{n}{k}$ is positive, which yields the first inequality stated. Because $\binom{n}{k} \leq \binom{n+1}{k}$, the second inequality follows. □

B.3 Minimum of Two Projections of an $\mathrm{N}_2(\mathbf{0}, \mathrm{I}_2)$ Random Vector

Lemma B.3.1. *Let $\boldsymbol{Z} = (Z_1, Z_2)^{\mathrm{T}}$ be distributed as $\mathrm{N}_2(\mathbf{0}, \mathrm{I}_2)$, i. e. bivariate normal with $\mathrm{E}[\boldsymbol{Z}] = \mathbf{0}$ and covariance matrix $\mathrm{I}_2 = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$. Further, let φ_1, φ_2 be the angles of the projections $Y_1 = (\cos\varphi_1, \sin\varphi_1)\boldsymbol{Z}$ and $Y_2 = (\cos\varphi_2, \sin\varphi_2)\boldsymbol{Z}$. Then the distribution of $\boldsymbol{Y} = (Y_1, Y_2)^{\mathrm{T}}$ is $\mathrm{N}_2(\mathbf{0}, \Sigma)$, with $\Sigma = \left(\begin{smallmatrix} 1 & \rho \\ \rho & 1 \end{smallmatrix}\right)$ and $\rho = \cos(\varphi_2 - \varphi_1)$.*

Proof. The expectations and variances of Y_1 and Y_2 are trivial. Let $\varphi = \varphi_2 - \varphi_1$. To calculate the covariance, assume, without loss of generality, that $\varphi_1 = 0$ (by a rotation of the coordinate system). Then $\operatorname{Cov}(Y_1, Y_2) = \operatorname{Cov}(Z_1, Z_1\cos\varphi + Z_2\sin\varphi) = \cos\varphi \cdot \operatorname{Var} Z_1 = \cos\varphi$. □

Lemma B.3.2. *Let $\boldsymbol{Y} = (Y_1, Y_2)^{\mathrm{T}}$ be distributed as $\mathrm{N}_2(\mathbf{0}, \Sigma)$ with $\Sigma = \left(\begin{smallmatrix} 1 & \rho \\ \rho & 1 \end{smallmatrix}\right)$. Then the density function of $Y_{\mathrm{m}} = \min(Y_1, Y_2)$ is*

$$f_{Y_{\mathrm{m}}}(y) = 2\phi(y)\left[1 - \Phi\left(y\sqrt{\frac{1-\rho}{1+\rho}}\right)\right],$$

where ϕ and Φ are the density function and the cumulative distribution function of the standard normal distribution, respectively.

Proof. Consider first $|\rho| < 1$. The joint density function of Y_1 and Y_2 is

$$\begin{aligned} f_{(Y_1,Y_2)}(y_1, y_2) &= \frac{1}{2\pi\sqrt{\det\Sigma}} \exp\left(-\frac{1}{2}\boldsymbol{y}^{\mathrm{T}}\Sigma^{-1}\boldsymbol{y}\right) \\ &= \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left(y_1^2 - 2\rho y_1 y_2 + y_2^2\right)\right). \end{aligned}$$

Integration over all (y_1, y_2) such that $\min(y_1, y_2) = y$ yields

$$\begin{aligned} f_{Y_{\mathrm{m}}}(y) &= \int_y^\infty f_{(Y_1,Y_2)}(y, t)\,dt + \int_y^\infty f_{(Y_1,Y_2)}(t, y)\,dt \\ &= 2\int_y^\infty \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left(y^2 - 2\rho y t + t^2\right)\right) dt \\ &= 2\,\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2 - y^2\rho^2}{2(1-\rho^2)}\right) \int_y^\infty \frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}} \exp\left(-\frac{(t-\rho y)^2}{2(1-\rho^2)}\right) dt \\ &= 2\phi(y)\left[1 - \Phi\left(y\sqrt{\frac{1-\rho}{1+\rho}}\right)\right]. \end{aligned}$$

For $\rho = 1$, the statement is obviously correct: $\Phi(0) = \frac{1}{2}$ leads to $f_{Y_{\mathrm{m}}}(y) = \phi(y)$. For $\rho = -1$, interpreting $[1 - (-1)]/[1 + (-1)]$ as $+\infty$ (because only the limit for

$\rho \downarrow -1$ is meaningful), and using $\Phi(+\infty) = 1$ and $\Phi(-\infty) = 0$, we can simplify the statement to

$$f_{Y_\mathrm{m}}(y) = \begin{cases} 2\phi(y) & y < 0, \\ 0 & y > 0, \end{cases}$$

which is also easily seen to be correct. □

Theorem B.3.3. *Let Y_1, Y_2 be projections of an $\mathrm{N}_2(\mathbf{0}, \mathrm{I}_2)$ random vector as in Lemma B.3.1. Then the density function of $Y_\mathrm{m} = \min(Y_1, Y_2)$ is*

$$f_{Y_\mathrm{m}}(y) = 2\phi(y) \left[1 - \Phi\left(y\sqrt{\frac{1-\rho}{1+\rho}}\right)\right],$$

where $\rho = \cos(\varphi)$ and φ is the angle between the two projections.

Proof. Immediate combination of the two preceding lemmas. □

B.4 Maximum of a Symmetric Binary Random Walk

Theorem B.4.1. *Let a symmetric binary random walk W_j on $\mathbb{Z}$ be defined as follows:*

$$\begin{aligned} \mathrm{P}(Z_i = -1) = \mathrm{P}(Z_i = +1) &= \frac{1}{2} \quad (i = 1, \ldots, n), \\ W_0 &= 0, \\ W_j &= \sum_{i=1}^{j} Z_i \quad (j = 1, \ldots, n), \end{aligned}$$

where the Z_i are independent.

Then, for $w \in \{-n, -n+2, \ldots, n-2, n\}$ and $b \in \{\max(0, w), \ldots, (n+w)/2\}$,

$$\mathrm{P}\left(\max_{0 \le j \le n} W_j \ge b \,\middle|\, W_n = w\right) = \frac{\binom{n}{(n+2b-w)/2}}{\binom{n}{(n+w)/2}}.$$

Proof. Note first that for each path starting at 0 and ending at w, the number of upward steps, $\sum_{i=1}^{n} 1(Z_i = +1)$, is $w + (n-w)/2 = (n+w)/2$.

We can use the reflection principle as follows to establish a one-to-one correspondence between paths starting at 0, reaching b at some time, and ending at w, and paths starting at 0 and ending at $2b - w$: From the first time j_0 the path

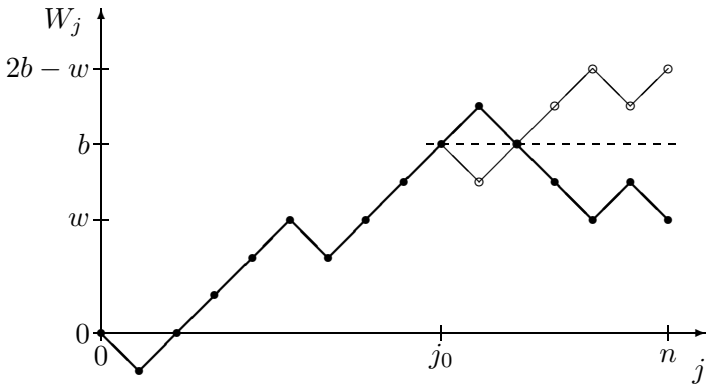

Figure B.1: A path starting at 0, reaching b at j_0, and ending at w (dots), and the corresponding path starting at 0 and ending at $2b-w$ (circles), which is obtained by reflecting the second part of the original path at the dashed line.

reaches b, we reflect the path at the horizontal line at level b (cf. Figure B.1). Because each such path (original or reflected) has probability $1/2^n$,

$$\begin{aligned} \mathrm{P}\left(\max_{0\le j\le n} W_j \ge b, W_n = w\right) &= \mathrm{P}(W_n = 2b-w) \\ &= \frac{1}{2^n}\binom{n}{(n+2b-w)/2}. \end{aligned}$$

The conditional probability is therefore

$$\begin{aligned} \mathrm{P}\left(\max_{0\le j\le n} W_j \ge b \,\middle|\, W_n = w\right) &= \frac{\mathrm{P}(\max_{0\le j\le n} W_j \ge b, W_n = w)}{\mathrm{P}(W_n = w)} \\ &= \frac{\frac{1}{2^n}\binom{n}{(n+2b-w)/2}}{\frac{1}{2^n}\binom{n}{(n+w)/2}} \\ &= \frac{\binom{n}{(n+2b-w)/2}}{\binom{n}{(n+w)/2}}. \end{aligned}$$

□

B.5 Convex Cones

Lemma B.5.1. *Let $C \subset \mathbb{R}^p$ be a convex cone, $\boldsymbol{a} \in C$, and $\boldsymbol{b} \in -(\mathbb{R}^p \setminus C)$. Then $\boldsymbol{a}+\boldsymbol{b} \in -(\mathbb{R}^p \setminus C)$.*

Proof. Assuming that $\boldsymbol{a} \in C$, $\boldsymbol{b} \in -(\mathbb{R}^p \setminus C)$, and that $\boldsymbol{a}+\boldsymbol{b} \notin -(\mathbb{R}^p \setminus C)$ leads to a contradiction: First note that $C + C \subset C$ because C is positively homogeneous and convex ($\boldsymbol{a}_1, \boldsymbol{a}_2 \in C \Rightarrow 2\boldsymbol{a}_1, 2\boldsymbol{a}_2 \in C \Rightarrow \boldsymbol{a}_1 + \boldsymbol{a}_2 = \frac{1}{2}(2\boldsymbol{a}_1) + \frac{1}{2}(2\boldsymbol{a}_2) \in C$).

$$\boldsymbol{a}+\boldsymbol{b} \notin -(\mathbb{R}^p \setminus C) \quad \Rightarrow \quad \boldsymbol{a}+\boldsymbol{b} \in -C \quad \Rightarrow \quad \boldsymbol{b} \in -C - C \subset -C,$$

which contradicts the second assumption. □

Appendix C

Algorithm for the Estimation of the Curve of Constant Power

The curves of constant power β in Figures 11.3–11.5 were estimated using the following algorithm:

1. Define a grid of points $\{-k\delta, -(k-1)\delta, \ldots, (k-1)\delta, k\delta\}^2$.

2. Estimate the power at $\boldsymbol{\vartheta} = \mathbf{0}$ by simulation.

3. If the estimated power is above β, move to the left and downwards, else move to the right and upwards on the diagonal. Estimate the power at each grid point by simulation; continue until β is crossed. Mark the last two points (one having estimated power above and one below β) visited on the diagonal as interesting.

4. For each interesting point P, estimate the power at its upper, lower, left, and right neighbors $N_1, \ldots, N_4$ on the grid by simulation. Mark all the neighbors as interesting if the set of the estimated power values at $P, N_1, \ldots, N_4$ contains values above and below β. Repeat this step until the power for the neighbors of all interesting points has been estimated. (This yields a band of points that should contain the desired curve.)

5. Drop all points with an estimated power that is too far away from β (e. g. using some binomial confidence interval based on the success probability β).

6. Use the principal curve (Hastie and Stuetzle, 1989) of the remaining points to estimate the curve of power β.

Steps 2.–4. are used to reduce computation time; in principle, we could simply estimate the power by simulation at all grid points instead.

If the curves are known to be symmetric about the diagonal $\vartheta_2 = \vartheta_1$ (as is the case for all examples in Figures 11.3 and 11.4), we only need to estimate the power for grid points on one side of the diagonal.

For the curves in Figures 11.3–11.5, a grid spacing of $\delta = 0.02$ and 5,000 simulations per grid point were used. For step 6., the R package princurve (S original by Trevor Hastie, R port by Andreas Weingessel), version 1.1-7, was used. The parameters for the principal.curve function were chosen as smoother='lowess' and f=0.05, which seemed to result in acceptably smooth curves without introducing too much bias.

Bibliography

Abelson, R. P., Tukey, J. W. (1963). Efficient utilization of non-numerical information in quantitative analysis: general theory and the case of simple order. *The Annals of Mathematical Statistics*, **34**, 1347–1369.

Ajne, B. (1968). A simple test for uniformity of a circular distribution. *Biometrika*, **55**, 343–354.

Akkerboom, J. C. (1990). *Testing problems with linear or angular inequality constraints*, Lecture Notes in Statistics, vol. 62, Springer, New York.

Bartholomew, D. J. (1961). A test of homogeneity of means under restricted alternatives. *Journal of the Royal Statistical Society, Series B*, **23**, 239–281.

Berger, R. L. (1982). Multiparameter hypothesis testing and acceptance sampling. *Technometrics*, **24**, 295–300.

Berger, R. L. (1989). Uniformly more powerful tests for hypotheses concerning linear inequalities and normal means. *Journal of the American Statistical Association*, **84**, 192–199.

Berger, R. L., Sinclair, D. F. (1984). Testing hypotheses concerning unions of linear subspaces. *Journal of the American Statistical Association*, **79**, 158–163.

Bhattacharyya, G. K., Johnson, R. A. (1970). A layer rank test for ordered bivariate alternatives. *The Annals of Mathematical Statistics*, **41**, 1296–1310.

Bloch, D. A., Lai, T. L., Tubert-Bitter, P. (2001). One-sided tests in clinical trials with multiple endpoints. *Biometrics*, **57**, 1039–1047.

Blumen, I. (1958). A new bivariate sign test. *Journal of the American Statistical Association*, **53**, 448–456.

Borovkov, A. A. (1998). *Mathematical statistics*, Gordon and Breach Science Publishers, Amsterdam.

Boyett, J. M., Shuster, J. J. (1977). Nonparametric one-sided tests in multivariate analysis with medical applications. *Journal of the American Statistical Association*, **72**, 665–668.

Bregenzer, T., Lehmacher, W. (1998). Directional tests for the analysis of clinical trials with multiple endpoints allowing for incomplete data. *Biometrical Journal*, **40**, 911–928.

Brown, B. M. (1983). Statistical uses of the spatial median. *Journal of the Royal Statistical Society, Series B*, **45**, 25–30.

Brown, B. M., Hettmansperger, T. P. (1987). Affine invariant rank methods in the bivariate location model. *Journal of the Royal Statistical Society, Series B*, **49**, 301–310.

Chakraborty, B., Chaudhuri, P., Oja, H. (1998). Operating transformation retransformation on spatial median and angle test. *Statistica Sinica*, **8**, 767–784.

Chatterjee, S. K. (1984). Restricted alternatives. In Krishnaiah, P. R., Sen, P. K. (eds.): *Handbook of statistics, vol. 4*, North-Holland, Amsterdam.

Chatterjee, S. K., De, N. K. (1972). Bivariate nonparametric location tests against restricted alternatives. *Calcutta Statistical Association Bulletin*, **21**, 1–20.

Chinchilli, V. M., Sen, P. K. (1981). Multivariate linear rank statistics and the union-intersection principle for the orthant restriction problem. *Sankhyā, Series B*, **43**, 152–171.

Choi, K., Marden, J. (1997). An approach to multivariate rank tests in multivariate analysis of variance. *Journal of the American Statistical Association*, **92**, 1581–1590.

Cohen, A., Sackrowitz, H. B. (1998). Directional tests for one-sided alternatives in multivariate models. *The Annals of Statistics*, **26**, 2321–2338.

Cohen, A., Sackrowitz, H. B. (2004). A discussion of some inference issues in order restricted models. *The Canadian Journal of Statistics*, **32**, 199–205.

Conaway, M. R., Petroni, G. R. (1996). Designs for phase II trials allowing for a trade-off between response and toxicity. *Biometrics*, **52**, 1375–1386.

De, N. (1976). Rank tests for randomized blocks against ordered alternatives. *Calcutta Statistical Association Bulletin*, **25**, 1–27.

Dietz, E. J. (1989). Multivariate generalizations of Jonckheere's test for ordered alternatives. *Communications in Statistics – Theory and Methods*, **18**, 3763–3783.

Dietz, E. J., Killeen, T. J. (1981). A nonparametric multivariate test for monotone trend with pharmaceutical applications. *Journal of the American Statistical Association*, **76**, 169–174.

Fahrmeir, L., Hamerle, A. (1984). *Multivariate statistische Verfahren*, de Gruyter, Berlin.

Fisher, R. A. (1932). *Statistical methods for research workers*, 4th ed., Oliver and Boyd, Edinburgh.

Follmann, D. (1996). A simple multivariate test for one-sided alternatives. *Journal of the American Statistical Association*, **91**, 854–861.

Freedman, L. S., Lowe, D., Macaskill, P. (1983). Stopping rules for clinical trials. *Statistics in Medicine*, **2**, 167–174.

Glimm, E., Srivastava, M. S., Läuter, J. (2002). Multivariate tests of normal mean vectors with restricted alternatives. *Communications in Statistics – Simulation and Computation*, **31**, 589–604.

Harville, D. A. (1997). *Matrix algebra from a statistician's perspective*, Springer, New York.

Hastie, T., Stuetzle, W. (1989). Principal curves. *Journal of the American Statistical Association*, **84**, 502–516.

Hedges, L. V., Olkin, I. (1985). *Statistical methods for meta-analysis*, Academic Press, Orlando.

Hettmansperger, T. P., McKean, J. W. (1998). *Robust nonparametric statistical methods*, Kendall's Library of Statistics, vol. 5, Arnold, London.

Hodges, J. L. (1955). A bivariate sign test. *The Annals of Mathematical Statistics*, **26**, 523–527.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, **6**, 65–70.

Hommel, G. (1983). Tests of the overall hypothesis for arbitrary dependence structures. *Biometrical Journal*, **25**, 423–430.

Hotelling, H. (1931). The generalization of Student's ratio. *The Annals of Mathematical Statistics*, **2**, 360–378.

Jennison, C., Turnbull, B. W. (1993). Group sequential tests for bivariate response: interim analyses of clinical trials with both efficacy and safety endpoints. *Biometrics*, **49**, 741–752.

Johnson, R. A., Mehrotra, K. G. (1972). Nonparametric tests for ordered alternatives in the bivariate case. *Journal of Multivariate Analysis*, **2**, 219–229.

Kost, J. T., McDermott, M. P. (2002). Combining dependent P-values. *Statistics & Probability Letters*, **60**, 183–190.

Kudô, A. (1963). A multivariate analogue of the one-sided test. *Biometrika*, **50**, 403–418.

Larocque, D., Labarre, M. (2004). A conditionally distribution-free multivariate sign test for one-sided alternatives. *Journal of the American Statistical Association*, **99**, 499–509.

Larocque, D., Tardif, S., van Eeden, C. (2000). Bivariate sign tests based on the sup, L_1 and L_2 norms. *Annals of the Institute of Statistical Mathematics*, **52**, 488–506.

Lehmann, E. L. (1952). Testing multiparameter hypotheses. *The Annals of Mathematical Statistics*, **23**, 541–552.

Logan, B. R. (2003). A cone order monotone test for the one-sided multivariate testing problem. *Statistics & Probability Letters*, **63**, 315–323.

Loughin, T. M. (2004). A systematic comparison of methods for combining p-values from independent tests. *Computational Statistics & Data Analysis*, **47**, 467–485.

Marcus, R., Peritz, E., Gabriel, K. R. (1976). On closed testing procedures with special reference to ordered analysis of variance. *Biometrika*, **63**, 655–660.

Mardia, K. V., Kent, J. T., Bibby, J. M. (1979). *Multivariate analysis*, Academic Press, London.

Maurer, W., Hothorn, L. A., Lehmacher, W. (1995). Multiple comparisons in drug clinical trials and preclinical assays: a-priori ordered hypotheses. In Vollmar, J. (ed.): *Biometrie in der chemisch-pharmazeutischen Industrie*, vol. 6, Fischer, Stuttgart.

McDermott, M. P., Wang, Y. (1999). Comment on "The emperor's new tests" by M. D. Perlman & L. Wu. *Statistical Science*, **14**, 374–377.

McDermott, M. P., Wang, Y. (2002). Construction of uniformly more powerful tests for hypotheses about linear inequalities. *Journal of Statistical Planning and Inference*, **107**, 207–217.

Merchant, J. A., Halprin, G. M., Hudson, A. R., Kilburn, K. H., McKenzie, W. N., Hurst, D. J., Bermazohn, P. (1975). Responses to cotton dust. *Archives of Environmental Health*, **30**, 222–229.

Minhajuddin, A. T. M., Frawley, W. H., Schucany, W. R., Woodward, W. A. (2006). Bootstrap tests for multivariate directional alternatives. *Preprint: http://www.smu.edu/statistics/TechReports/TR333.pdf.*

Möttönen, J., Hüsler, J., Oja, H. (2003). Multivariate nonparametric tests in a randomized complete block design. *Journal of Multivariate Analysis*, **85**, 106–129.

Möttönen, J., Oja, H. (1995). Multivariate spatial sign and rank methods. *Nonparametric Statistics*, **5**, 201–213.

Möttönen, J., Oja, H., Tienari, J. (1997). On the efficiency of multivariate spatial sign and rank tests. *The Annals of Statistics*, **25**, 542–552.

Mudholkar, G. S., Kost, J., Subbaiah, P. (2001). Robust tests for the significance of orthant restricted mean vector. *Communications in Statistics – Theory and Methods*, **30**, 1789–1810.

Neuhaus, G., Zhu, L.-X. (1999). Permutation tests for multivariate location problems. *Journal of Multivariate Analysis*, **69**, 167–192.

O'Brien, P. C. (1984). Procedures for comparing samples with multiple endpoints. *Biometrics*, **40**, 1079–1087.

Oja, H. (1999). Affine invariant multivariate sign and rank tests and corresponding estimates: a review. *Scandinavian Journal of Statistics*, **26**, 319–343.

Park, H.-I., Na, J.-H., Desu, M. M. (2001). Nonparametric one-sided tests for multivariate data. *Sankhyā, Series B*, **63**, 286–297.

Perlman, M. D. (1969). One-sided testing problems in multivariate analysis. *The Annals of Mathematical Statistics*, **40**, 549–567.

Perlman, M. D., Chaudhuri, S. (2004). The role of reversals in order-restricted inference. *The Canadian Journal of Statistics*, **32**, 193–198.

Perlman, M. D., Wu, L. (2006). Some improved tests for multivariate one-sided hypotheses. *Metrika*, **64**, 23–39.

Pocock, S. J., Geller, N. L., Tsiatis, A. A. (1987). The analysis of multiple endpoints in clinical trials. *Biometrics*, **43**, 487–498.

Randles, R. H. (1989). A distribution-free multivariate sign test based on interdirections. *Journal of the American Statistical Association*, **84**, 1045–1050.

Randles, R. H. (2000). A simpler, affine-invariant, multivariate, distribution-free sign test. *Journal of the American Statistical Association*, **95**, 1263–1268.

Reitmeir, P., Wassmer, G. (1996). One-sided multiple endpoint testing in two-sample comparisons. *Communications in Statistics – Simulation and Computation*, **25**, 99–117.

Robertson, T., Wright, F. T., Dykstra, R. L. (1988). *Order restricted statistical inference*, Wiley, Chichester.

Röhmel, J., Streitberg, B. (1987). Zur Konstruktion globaler Tests. *EDV in Medizin und Biologie*, **18**, 7–11.

Roy, S. N. (1953). On a heuristic method of test construction and its use in multivariate analysis. *The Annals of Mathematical Statistics*, **24**, 220–238.

Rüger, B. (1978). Das maximale Signifikanzniveau des Tests: „Lehne H_0 ab, wenn k unter n gegebenen Tests zur Ablehnung führen“. *Metrika*, **25**, 171–178.

Sasabuchi, S. (1988). A multivariate one-sided test with composite hypotheses when the covariance matrix is completely unknown. *Memoirs of the Faculty of Science, Kyushu University, Series A*, **42**, 37–46.

Sen, P. K., Silvapulle, M. J. (2002). An appraisal of some aspects of statistical inference under inequality constraints. *Journal of Statistical Planning and Inference*, **107**, 3–43.

Shapiro, A. (1988). Towards a unified theory of inequality constrained testing in multivariate analysis. *International Statistical Review*, **56**, 49–62.

Silvapulle, M. J., Silvapulle, P., Basawa, I. V. (2002). Tests against inequality constraints in semiparametric models. *Journal of Statistical Planning and Inference*, **107**, 307–320.

Small, C. G. (1990). A survey of multidimensional medians. *International Statistical Review*, **58**, 263–277.

Tamhane, A. C., Logan, B. R. (2004). A superiority-equivalence approach to one-sided tests on multiple endpoints in clinical trials. *Biometrika*, **91**, 715–727.

Tang, D.-I. (1998). Testing the hypothesis of a normal mean lying outside a convex cone. *Communications in Statistics – Theory and Methods*, **27**, 1517–1534.

Tang, D.-I., Gnecco, C., Geller, N. L. (1989). An approximate likelihood ratio test for a normal mean vector with nonnegative components with application to clinical trials. *Biometrika*, **76**, 577–583.

Tippett, L. H. C. (1937). *The methods of statistics*, 2nd ed., Williams and Norgate, London.

Tsai, K.-T., Koziol, J. A. (1994). Assessing multiple endpoints with ordered alternatives. *Communications in Statistics – Theory and Methods*, **23**, 533–544.

Tsai, M.-T. M., Sen, P. K. (1990). Asymptotically efficient rank MANOVA tests for restricted alternatives in randomized block designs. *Annals of the Institute of Statistical Mathematics*, **42**, 375–385.

Tyler, D. E. (1987). A distribution-free M-estimator of multivariate scatter. *The Annals of Statistics*, **15**, 234–251.

van der Vaart, A. W. (1998). *Asymptotic statistics*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, Cambridge.

Vock, M. (2006). Graphical comparison of multivariate nonparametric location tests for restricted alternatives. *Statistics & Probability Letters*, **76**, 1529–1535.

Wei, L. J., Knuiman, M. W. (1987). A one-sided rank test for multivariate censored data. *The Australian Journal of Statistics*, **29**, 214–219.

Wei, L. J., Lin, D. Y., Weissfeld, L. (1989). Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *Journal of the American Statistical Association*, **84**, 1065–1073.

Wellek, S. (2003). *Testing statistical hypotheses of equivalence*, Chapman & Hall/CRC, Boca Raton.

Westfall, P. H., Young, S. S. (1993). *Resampling-based multiple testing: examples and methods for p-value adjustment*, Wiley, New York.

Witting, H. (1985). *Mathematische Statistik I*, Teubner, Stuttgart.

Witting, H., Nölle, G. (1970). *Angewandte mathematische Statistik*, Teubner, Stuttgart.

Index

acceptance region, 51, 74
affine transformation, 11
angle test, 19
asymptotics, 101

binomial distribution, 29–30, 102

component permutation, 11
componentwise homeomorphism, 12
componentwise strictly monotone transformation, 12
composite hypothesis, 7
cone, 5
  convex, 5, 105
  s-simply-connected, 7
cone order monotonicity, 15, 79–80
cone/opposite cone test, 29–30
confidence region, 73
curve of constant power, 58–62, 106

diagonal symmetry, 17, 96
direction alternative, 4
directional symmetry, 17, 96
distance
  of sets, 77
dual
  positive, *see* positive dual

elliptical directions class, 16
elliptical symmetry, 16
exponential radius distribution, 71

Hausdorff distance, 77
Hodges(–Ajne) test, 44
homeomorphism, 12

indifference region, 8
inequality symbols
  for vectors, 19
infimum distance, 77
intersection-union test, 28
invariance, 12

location transformation, 11

meta-analysis, 25
min test, 28, 76
monotone trend problem, 10, 24
monotonicity
  cone order, *see* cone order monotonicity
multi-sample problem, 10, 23
multiple testing, 24, 26, 83

non-inferiority, 8, 66, 69

O'Brien/Wilcoxon test, 63
one-quadrant alternative, 5
one-sample problem, 3, 19
orthant, 7
orthogonal transformation, 11

point hypothesis, 7
positive dual, 15, 46
positively homogeneous set, 5
principal curve, 106
punched Hausdorff distance, 77

random walk
  binary, symmetric, 47, 104
rank
  spatial, *see* spatial rank
reflection principle, 104
rejection region, 51

rotation, 11

scale transformation, 11
sector alternative, 5
sign
    spatial, *see* spatial sign
simple hypothesis, 7
spatial rank, 32
spatial sign, 32, 96–100
spatial sign test, 32–38
spatial signed rank test, 38
spherical directions class, 16
spherical symmetry, 16
strongly stochastically smaller, 22
symmetry, 16, 96

t distribution, 69
three-quadrant alternative, 5
transformation, 11–14
translation, 11
    of hypotheses, 10
two-sample problem, 10, 22

unbiasedness
    confidence region, 74–75
    test, 14, 58, 75
union-intersection test, 28

Lebenslauf

Personalien	Michael Vock geboren am 10. Januar 1977 in Bern Heimatorte: Aarau und Wohlen (AG)	
Bildungsweg	1984–1990	Primar- und Sekundarschule in Spiegel (BE)
	1990–1997	Untergymnasium und Gymnasium (Typus B) in Köniz (BE)
	1997–2002	Studium in mathematischer Statistik und Versicherungslehre an der Universität Bern Nebenfächer: Mathematik, Informatik Diplomarbeit: *Ranked Set Sampling: Praktische Probleme und deren Erkennung*, unter der Leitung von Prof. Dr. J. Hüsler
	2002–2007	Doktorat an der Universität Bern
Anstellungen	2000–2002	Hilfsassistent am IMSV der Universität Bern
	2002–	Assistent am IMSV der Universität Bern
	2006	Wissenschaftlicher Mitarbeiter bei ISS AG, Port