Natural bootstrap method to test for some symmetries of a multivariate distribution.

Trabajo presentado por David Pardo
para obtener el grado de Magister en Matemáticas

Asesor: Adolfo Quiroz

Universidad de los Andes

Facultad de Ciencias

Departamento de Matemáticas

June 1, 2014

Acknowledgements

I would like to express my deep gratitude to Adolfo José Quiroz, my supervisor in this project, for his guidance, encouragement and constant help. I also wish to thank the Department of Mathematics for its support through the program of graduate assistants as well as to all the people involved in my postgraduate education. Finally, my grateful thanks to my friends and family, whom were inconditionally by my side during this process, specially to my dad Campo Elías, my mom Esperanza, my brother Camilo and my girlfriend Tata.

Contents

Introduction

The objective of the present work is to develop a new bootstrap method for testing for some symmetries of a multivariate distribution, based on previous results by Koltchinskii and Sakhanenko involving elliptical symmetry. Along the path leading to this goal we discuss the concept of symmetry from a statistical perspective, we delve into some results from the field of empirical processes and we study the details of the method presented in Koltchinskii and Sakhanenko (2000).

In Chapter 1 we give an overview of the different concepts of multivariate symmetry of a distribution, focusing in spherical, elliptical, central and angular symmetry. For these cases we give an overview of some results concerning to the problem of testing for these symmetries. We finish by mentioning some other definitions of symmetry which generalize the concept in different directions.

We continue to the study of empirical processes in Chapter 2, looking forward to presenting uniform extensions of the Law of Large Numbers and the Central Limit Theorem. In the process we introduce Glivenko-Cantelli and Donsker classes of functions and the basics of Vapnik-Čerbonenkis theory.

With the theory developed in the previous chapters we are ready to tackle, in Chapter 3, the method of Koltchinskii and Sakhanenko for testing for elliptical symmetry of a multivariate distributions. We give the details of the method without proving their results of convergence since we deal with them in the next chapter.

Chapter 4 is the main part of this work, where we develop a method for testing for central symmetry giving all the details and the most important proofs of convergence. We discuss the specifics of how the method can be implemented and we extend it to angular symmetry. We finish with suggestions for applying the method to other kinds of symmetry.

Chapter 5 is concerned with the results of the implementation of the method and the analysis of its behaviour.

At the end we present our conclusions and prospects for further work.

Chapter 1

Symmetry

Symmetry is a property which arises naturally and is found everywhere, from the tiniest of atoms to the large scale structure of the universe. In particular, it is an important feature of lots of the data produced in human endeavors. When a statistical analysis is to be made on data, knowledge about its symmetry properties could be essential to determine the validity of certain processes or it could allow the use of tools developed for data with specific characteristics. Examples of the use of symmetry for statistical analysis are abundant; a short overview of applications in a wide range of areas, from finance to genetics, can be consulted in Sakhanenko (2009). A detailed account of the use of symmetry in scientific fields is given in Petitjean (2003) and an interesting example of the need of studying symmetry of distributions for policy making is presented in Zheng and Gastwirth (2010).

Due to its theoretical importance and its practical applications, the problem of finding symmetry properties of a data set or of its underlying distribution is of great interest. The first step to tackle this problem is to give a precise definition of symmetry. In the univariate case this is straightforward:

Definition 1.1. A random variable X is *symmetric* around a point θ if $X-\theta \stackrel{d}{=} \theta-X$. In terms of the density function f, this becomes $f(\theta - x) = f(\theta + x)$.

The problem of testing for symmetry of a univariate distribution, i. e. to develop tests that can identify whether a given univariate distribution or data set is symmetric, has been widely studied. Some important references are given in Shorack and Wellner (1986) and a bootstrap method related with the present work is developed in Arcones and Giné (1991).

The extension of the concept of symmetry for multivariate distributions is not unique. A beautiful summary of the main definitions of multivariate symmetry is presented in Serfling (2006). We will focus in the four basic extensions of univariate symmetry: *spherical, elliptical, central* and *angular symmetry.*

1.1 Spherical symmetry

Definition 1.2. A random vector $\mathbf{X}$ has a *spherically symmetric* distribution about θ if

$$\mathbf{X} - \boldsymbol{\theta} \stackrel{d}{=} \mathbf{A}(\mathbf{X} - \boldsymbol{\theta})$$

for all orthogonal $d \times d$ matrices $\mathbf{A}$. Here $\stackrel{d}{=}$ denotes convergence in distribution.

Characteristic function. The characteristic function of a spherically symmetric distribution is given by $e^{i\mathbf{t}^T\boldsymbol{\theta}}h(\mathbf{t}^T\mathbf{t})$, for some scalar function h.

Density. If the distribution has a density, it has the form $g((\mathbf{x} - \boldsymbol{\theta})^T(\mathbf{x} - \boldsymbol{\theta}))$, $\mathbf{x} \in \mathbb{R}^d$, for some nonnegative scalar function g.

Geometry. Geometrically, the definition means that the distribution is invariant under rotations. This also accounts for the contours of equal density to be spheres in $\mathbb{R}^{d-1}$.

Examples. The most common example of a spherically symmetrical distribution is a multivariate normal distribution with covariance matrix given by $\sigma^2\mathbf{I}_d$. A trivial, however important, example is the uniform distribution over the unitary sphere in

d dimensions, $\mathbb{S}^{d-1}$. Other distributions with this property are certain cases of standard multivariate t and logistic distributions, for details consult Kotz et al. (2000).

Other characterizations. An example particularly useful for testing for symmetry is that the radial random vectors $\|\mathbf{X} - \theta\|$ and the corresponding random directional unitary vector $(\mathbf{X} - \boldsymbol{\theta})/\|\mathbf{X} - \boldsymbol{\theta}\|$ be independent with this last random vector having uniform distribution over $\mathbb{S}^{d-1}$ (see Dempster (1969)).

Another chacarterization is that the projections of $\mathbf{X} - \boldsymbol{\theta}$ onto lines through the origin have identical univariate distributions. However, this kind of characterizations is not as interesting for testing for symmetry, because they add an additional step, where randomness is involved, for choosing the projections to be considered.

Testing for symmetry. Testing the hypothesis of symmetry of a multivariate distribution is still a research problem nowadays and several approaches to tackle it have been developed. Using the first characterization for spherical symmetry above, uniform most powerful tests have been proposed by Kariya and Eaton (1977) and Gupta and Varga (1993); and a bivariate Cramer-von Mises type test has been extended to multivariate spherical symmetry in Baringhaus (1991); Zhu et al. (1995) presented a test based on the projection onto lines characterization of this symmetry and graphical methods were developed by Li et al. (2001). Yet another approach investigated in Koltchinskii and Li (2000) is to define measures of asymmetry and apply them on the sample to obtain a statistic. Bootstrap tests are given in Romano (1989) and Monte Carlo methods are presented in Zhu and Neuhaus (2000).

There are some considerations to take into account when selecting a particular symmetry test method. It is desired to have a computationally efficient test with a wide range of application and good power against alternative hypothesis. The choice of the test would depend in an equilibrium of these three factors, e.g. uniformly most powerful test are the best possible selection if power is the most relevant feature; however they are commonly restricted to a narrow family of distributions. On the other hand nonparametric bootstrap and Monte Carlo approaches can be used to test several alternative hypothesis but at the cost of power and computational time. In conclusion, there is no superior method for testing for symmetry, since the particular application in hand dictates the needs for certain features of the test, thus the selection must be made casewise; moreover, this encourages the study of new methods which may optimize features not considered in the already existing tests.

1.2 Elliptical symmetry

Although the family of spherical symmetric distributions have many desirable features for statistical analyses, this symmetry encompasses too restricting conditions. Elliptical symmetry is a straightforward generalization which preserves most of the important characteristics of the spherically symmetric distributions, while admiting a considerable number of new distributions. A general overview of this family of distributions is found in Anderson and Fang (1990).

Definition 1.3. A random vector $\mathbf{X}$ has an *elliptically symmetric* distribution with parameters $\boldsymbol{\theta}$ and $\boldsymbol{\Sigma}$ if there exists a spherically symmetric random vector $\mathbf{Y}$ such that

$$\mathbf{X} \stackrel{d}{=} \mathbf{A}^T\mathbf{Y} + \boldsymbol{\theta},$$

where $\mathbf{A}_{k\times d}$ satisfies $\mathbf{A^TA} = \boldsymbol{\Sigma}$ with rank $\boldsymbol{\Sigma} = k \leq d$.

Characteristic function. The characteristic function of an elliptically symmetric distribution has the form $e^{i\mathbf{t}^T\boldsymbol{\theta}}h(\mathbf{t}^T\boldsymbol{\Sigma}\mathbf{t})$, for some scalar function h.

Density. If the distribution has a density, it has the form $|\boldsymbol{\Sigma}|^{-1/2}g((\mathbf{x}-\boldsymbol{\theta})^T\boldsymbol{\Sigma}^{-\mathbf{1}}(\mathbf{x}-\boldsymbol{\theta}))$, $\mathbf{x} \in \mathbb{R}^d$, for some nonnegative scalar function g. If $\mathbf{A}$ is nonsingular, the density may be represented by $|\mathbf{A}|^{-1}g_0(\mathbf{A}^{-\mathbf{1}}(\mathbf{x}-\boldsymbol{\theta}))$, for a density g_0 spherically symmetric about the origin.

Geometry. The definition says that the distribution $\mathbf{X}$ is affinely equivalent to that of a spherically symmetric random vector $\mathbf{Y}$, thus the family of elliptically symmetric distributions is closed under affine transformations. We also have that the contours of equal density has elliptical form.

Examples. Multivariate normal distributions $N(\theta, \mathbf{\Sigma})$ and multivariate t-distributions $T(m, \theta, \mathbf{\Sigma})$ are the main examples of families of elliptically symmetric distributions.

Other characterizations. Since an ellipticallly symmetric distribution is an affine transformation of a spherically symmetric distribution, in applications it is possible to apply the inverse transformation and use the alternative characterizations of spherical symmetry given in the previous section.

Testing for symmetry. A method based in the representation of the nonsigular $\mathbf{A}$ is illustrated in Beran (1979). Its test statistic is defined in terms of the directions and distances of the residuals $\hat{\mathbf{A}}_n^{-1}(X_i - \hat{\boldsymbol{\theta}}_n)$ with respect to the origin, where sample estimates $\hat{\boldsymbol{\theta}}_n$ and $\hat{\mathbf{A}}_n$ are used. In Manzotti et al. (2002) a method based in the characterization of the spherical symmetry as a decomposition in radial and angular distributions is introduced. They average spherical harmonics over scaled residuals of the distribution, which are equivalent to the residuals in the previous method but restricted to the unitary sphere. A bootstrap method based on functionals of empirical processes indexed by certain classes of functions is found in Koltchinskii and Sakhanenko (2000). Finally, Monte Carlo, projection and graphical methods are presented in Zhu and Neuhaus (2000), Fang et al. (1995) and Li et al. (2001) respectively.

1.3 Central and angular symmetry

The most direct generalization of univariate symmetry is central symmetry, also called diagonal or reflective symmetry. It is a more general concept of symmetry than those previously defined. An even broader generalization, directly related with central symmetry is angular symmetry. Let us define both.

Definition 1.4. A random vector $\mathbf{X}$ has a *centrally symmetric* distribution about $\boldsymbol{\theta}$ if

$$\mathbf{X} - \boldsymbol{\theta} \stackrel{d}{=} (\boldsymbol{\theta} - \mathbf{X}).$$

Definition 1.5. A random vector $\mathbf{X}$ has an *angularly symmetric* distribution about θ if

$$\frac{\mathbf{X} - \boldsymbol{\theta}}{\|\mathbf{X} - \boldsymbol{\theta}\|} \stackrel{d}{=} \frac{(\boldsymbol{\theta} - \mathbf{X})}{\|\mathbf{X} - \boldsymbol{\theta}\|}.$$

Observe that an equivalent definition of angular symmetry is that the random vector $\frac{\mathbf{X} - \boldsymbol{\theta}}{\|\mathbf{X} - \boldsymbol{\theta}\|}$ has a centrally symmetric distribution. Thus, the properties of central symmetry can be extrapolated to angular symmetry for the mentioned random vector.

Although the generality of these symmetries results in a lesser range of applications, they are still of great importance. A simple yet ubiquous example is that of trimming; if a dataset is to be trimmed symmetrically in each marginal, i.e. the first and last 5% of data in each coordinate are to be removed, one has to check that the data have central symmetry for the trimming to be statistically valid.

Density. If the density exists, it satisfies $f(\boldsymbol{\theta} - \mathbf{x}) = f(\mathbf{x} - \boldsymbol{\theta})$.

Geometry. A distribution with central symmetry is invariant under reflections with respect to a point. The contours of equal density also have this property. For angular symmetry, projections onto the unit sphere have these properties.

Examples. Central symmetric functions includes all elliptically symmetric distributions. Other examples are the uniform distributions over reflectively symmetric polygons, such as squares. Angular symmetry encompasses this family; a distribution which is angularly symmetric but not centrally symmetric is one which can be expressed as $P = CR$, a product of a centrally symmetric distribution C and a radial distribution R that varies depending on the direction. Such kind of distribution will have centrally symmetric projections onto the unitary sphere, but the changing radii do not allow the central symmetry of the whole distribution.

Other characterizations. An equivalent definition for central symmetry is that all the projection of $\mathbf{X} - \boldsymbol{\theta}$ onto lines through the origin have symmetric univariate distributions. A distribution is angularly symmetric iff any hyperplane passing through $\boldsymbol{\theta}$ divides $\mathbb{R}^d$ into two open half-spaces with equal probabilities (one half each).

Testing for symmetry. Tests for central symmetry can be adapted to test for angular symmetry, thus we do not specify which of the symmetries are being tested in each case. In Blough (1989) and Ghosh and Ruymgaart (1992) projection methods are used, taking advantage of the alternative characterizations involving projection onto lines through the origin. In the latter the empirical characteristic function is used. Monte Carlo methods are developed in Dicks and Tong (1999) and Zhu and Neuhaus (2000), and graphical methods are shown in Liu et al. (1999). An extension of the bootstrap test for elliptical symmetry of Koltchinskii and Sakhanenko (2000) is presented in Sakhanenko (2009), where central and angular symmetry are shown as examples of the broader concept of group symmetry, that is defined bellow.

1.4 Other definitions of symmetry

We now present some generalizations of the symmetries studied above. We begin with the most relevant generalization in our pursuit, the so named group symmetry, which encloses the four symmetries that were presented in detail. Let $\mathcal{S}$ be a compact group of linear transformations from $\mathbb{R}^d$ to $\mathbb{R}^d$.

Definition 1.6. A random vector $\mathbf{X}$ has a $\mathcal{S}$-*symmetric* distribution iff there exists an affine nonsigular transformation $\mathbf{A}_0$ such that the random vector $\mathbf{Z} = \mathbf{A}_0\mathbf{X}$ is $\mathcal{S}$-invariant, i.e.

$$\mathbf{Z} \stackrel{d}{=} \mathbf{A}_0\mathbf{X},$$

for all transformations $S \in \mathcal{S}$.

Notice that this definition reduces to the case of spherical symmetry when $\mathcal{S}$ is the group of all orthogonal transformations in $\mathbb{R}^d$. Similarly, ellyptical symmetry is found when the transformations satisfies $\mathbf{A}_0(\mathbf{x}) = \mathbf{V}_0^{-1}(\mathbf{x} - \boldsymbol{\theta})$, where V_0 is the square root of the covariance operator of the random vector $\mathbf{X}$. Central symmetry is also included by taking $\mathcal{S}$ consisting of the identity transformation and its negative.

Definition 1.7. A random vector $\mathbf{X}$ has a *sign-symmetric* distribution about $\boldsymbol{\theta}$ if

$$\mathbf{X} - \boldsymbol{\theta} = (X_1 - \theta_1, ..., X_d - \theta_d) \stackrel{d}{=} (\pm(X_1 - \theta_1), ..., \pm(X_d - \theta_d)),$$

for all combinations of $-, +$.

This symmetry is more restrictive than central symmetry and yet includes examples such as uniform distributions over squares. It can be represented in terms of $\mathcal{A}_0-$symmetry for $\mathcal{A}_0$ consisting of the 2^d transformations given by vectors of the form $(\pm 1, ..., \pm 1)$ for all choices of $-, +$.

A generalization of angular symmetry called half-space symmetry is presented in Zuo and Serfling (2000).

Definition 1.8. A random vector $\mathbf{X}$ has a *half-space symmetric* distribution about $\boldsymbol{\theta}$ if

$$P(\mathbf{X} \in H) \geq 1/2, \text{for every half-space } H \text{ with } \boldsymbol{\theta} \text{ on the boundary.}$$

This definition is equivalent to angular symmetry except for discrete distributions for which the point $\boldsymbol{\theta}$ has positive probability.

Chapter 2

Empirical Processes

In this chapter we will give the theoretical foundations of the symmetry test that is proposed. The main objective is to study the convergence of the empirical process for certain classes of functions. The present approach is based on the account of the subject given on van der Vaart and Wellner (1996).

2.1 Introduction

Definition 2.1. Let $(X_1, ..., X_n)$ be an independent identically distributed (from now on this will be called i.i.d.) random sample in a measurable space $(\mathcal{X}, \mathcal{A})$. The *empirical measure* $\mathbb{P}_n$ of this sample is the discrete random measure defined by $\mathbb{P}_n(A) = n^{-1}\sum_{i=1}^n I_A(X_i) = n^{-1}\sum_{i=1}^n \delta_{X_i}(A)$, where I_A is the indicator function and δ_{X_i} is the Dirac measure, i.e. $\delta_{X_i} =$

Let Q be a signed measure (allowing it to have negative values) and f a measurable function of a class $\mathcal{F}$. We will use the notation $Qf := \int f dQ$ and $\|Q\|_{\mathcal{F}} := \sup\{|Qf| : f \in \mathcal{F}\}$.

Definition 2.2. Let P be the underlying distribution of the X_i and $\mathcal{F}$ a collection of measurable functions $f : \mathcal{X} \to \mathbb{R}$. The $\mathcal{F}$-indexed *empirical process* $\mathbb{G}_n$ is defined by

$$\mathbb{G}_n f := \sqrt{n}(\mathbb{P}_n - P)f = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (f(X_i) - Pf)$$

From the law of large numbers and the central limit theorem we have for a given function f

$$\mathbb{P}_n f \xrightarrow{a.s.} Pf,$$

$$\mathbb{G}_n f \rightsquigarrow N(0, P(f - Pf)^2),$$

provided that Pf exists and $Pf^2 < \infty$ respectively. Here $\rightsquigarrow$ denotes weak convergence and $a.s.$ stands for almost surely.

We want to find conditions for which the above results hold uniformly over a class of functions. For the law of large numbers this means

$$\|\mathbb{P}_n - P\|_{\mathcal{F}} \longrightarrow 0, \tag{2.1}$$

where the convergence is in outer probability or is outer almost surely.

Definition 2.3. A class of functions $\mathcal{F}$ is called a *P-Glivenko-Cantelli class* if 2.1 holds over $\mathcal{F}$.

For the central limit theorem, assuming that

$$\sup_{f \in \mathcal{F}} |f(x) - Pf| < \infty, \text{for every } x,$$

the result becomes

$$\mathbb{G}_n = \sqrt{n}(\mathbb{P}_n - P) \rightsquigarrow \mathbb{G}, \text{ in } l^\infty(\mathcal{F}), \tag{2.2}$$

where $\mathbb{G}_n := n^{-1/2} \sum_{i=1}^{n} (\delta_{X_i} - P)$ is the signed measure identified with the empirical process and $\mathbb{G}$ is a tight Borel measurable element in $l^\infty(\mathcal{F})$.

To understand the limit process we analyze the marginal distributions. From the multivariate central limit theorem, we have that for any finite set $f_1, ..., f_k$ of functions

$$(\mathbb{G}_n f_1, ..., \mathbb{G}_n f_k) \rightsquigarrow N_k(\mathbf{0}, \mathbf{\Sigma})$$

iff the functions are square integrable. Here $\mathbf{\Sigma}$ is a $k \times k$-matrix with elements (i, j) given by $P(f_i - Pf_i)(f_j - Pf_j)$. Thus, taking into account that convergence in $l^\infty(\mathcal{F})$ implies marginal convergence, we obtain that the empirical process converges to a zero-mean Gaussian process with covariance function

$$E\mathbb{G}f_1\mathbb{G}f_2 = P(f_1 - Pf_1)(f_2 - Pf_2) = Pf_1f_2 - Pf_1Pf_2.$$

These properties and tightness completely determine the limit process $\mathbb{G}$, which is called the *P-Brownian bridge.*

Definition 2.4. A class of functions $\mathcal{F}$ is called a *P-Donsker class* if 2.2 holds over $\mathcal{F}$.

2.2 Glivenko-Cantelli and Donsker theorems

We want to identify P-Glivenko-Cantelli and P-Donsker classes of functions. These properties are directly related with the size of the classes, therefore we first define a way to measure it.

Definition 2.5. Given Q a probability measure and $\mathcal{F}$ a class of functions in $L^r(Q)$, for each $\epsilon > 0$ the *covering number* $N(\epsilon, \mathcal{F}, L^r(Q))$ is the smallest value of m for which there exist functions $g_1, ..., g_m$ such that $\min_j P|f - g_j| \leq \epsilon$ for each $f \in \mathcal{F}$. If no such m exists, $N(\epsilon, \mathcal{F}, L^r(Q)) = \infty$. The *entropy* is the logarithm of the covering number.

Informally, the covering number is the smallest number of balls of radius ϵ needed to cover $\mathcal{F}$. Evidently, this number gives a notion of the size of a class; however, since it depends on ϵ, we will be interested in its rate of increment with decreasing ϵ rather than in the number for some fixed radius.

Definition 2.6. An *envelope function* of a class $\mathcal{F}$ is any function F such that for every $f \in \mathcal{F}$ and every x, $|f(x)| \leq |F(x)|$.

Definition 2.7. The *uniform entropy numbers* relative to L^r are given by

$$\sup_Q \log N(\epsilon \|F\|_{Q,r}, \mathcal{F}, L^r(Q)),$$

where the supremum is taken over over all probability measures Q on $(\mathcal{X}, \mathcal{A})$, with $0 < QF^r < \infty$.

Definition 2.8. A class $\mathcal{F}$ of measurable functions on a probability space $(\mathcal{X}, \mathcal{A}, P)$ is *P-measurable* if the function $(X_1, ..., X_n) \mapsto \| \sum_{i=1}^n e_i f(X_i) \|_{\mathcal{F}}$ is measurable on the completion of $(\mathcal{X}^n, \mathcal{A}^n, P^n)$ for every n and every vector $(e_1, ..., e_n) \in \mathbb{R}^n$.

In applications it is hard to find classes of function which are not $P-$measurable, thus this measurability requirement in the following theorems is not as strong as it could seem.

We now can state the main results relating covering numbers with uniform convergence of empirical processes, the Glivenko-Cantelli and Donsker theorems. The Glivenko-Cantelli theorem will give us an empirical law of large numbers, whereas the Donker theorem will provide an empirical central limit theorem.

Theorem 2.9. *(Glivenko-Cantelli) Let $\mathcal{F}$ be a $P-$ measurable class of measurable functions with envelope F such that $P^*F < \infty$. Let $\mathcal{F}_M$ be the class of functions $fI_{F \leq M}$ when f ranges over $\mathcal{F}$. If $\log N(\epsilon, \mathcal{F}_M, L^1(\mathbb{P}_n)) = o_P^*(n)$ for every ϵ and $M > 0$, then $\|\mathbb{P}_n - P\|_{\mathcal{F}}^* \to 0$ both almost surely and in mean. In particular, $\mathcal{F}$ is $P-$Glivenko-Cantelli.*

Definition 2.10. The *uniform entropy bound* is defined as

$$\int_0^\infty \sup_Q \sqrt{\log N(\epsilon\|F\|_{Q,2}, \mathcal{F}, L^2(Q))}d\epsilon < \infty, \tag{2.3}$$

with $\|F\|_{Q,2} = (\int |F|^2 dQ)^{1/2}$.

Theorem 2.11. *(Donsker) Let $\mathcal{F}$ be a class of measurable functions that satisfies the uniform entropy bound 2.3. Let the classes $\mathcal{F}_\delta = \{f-g : f, g \in \mathcal{F}, \|f-g\|_{P,2} < \delta\}$ and $\mathcal{F}^2_\infty$ be $P-$measurable for every $\delta > 0$. If $P^*F^2 < \infty$, then $\mathcal{F}$ is $P-$Donsker.*

The proofs of these results are somewhat involved and require previous results which are not directly related with the topic of this work, hence we will not present them here. Both proofs are found in van der Vaart and Wellner (1996) and a very detailed account of the Glivenko-Canelli Theorem is presented in Pollard (1984). We finish this section with one important fact about Donsker classes.

Definition 2.12. An empirical process $\mathbb{G}_n$ on $\mathcal{F}$is *asymptotically equicontinuous* if for every $\epsilon > 0$, it satisfies

$$\lim_{\delta\to 0} \limsup_{n\to\infty} P^*(\sup_{\rho_P(f-g)<\delta} |\mathbb{G}_n(f-g)| > \epsilon) = 0,$$

where ρ_P is a seminorm defined by $\rho_P(f) := (P(f - Pf)^2)^{1/2}$.

This notion of asymptotical equicontinuity will be used extensively in proving the convergence of the statistic used in the proposed test for symmetry. The following theorem is the foundation for its usefulness.

Theorem 2.13. *A class $\mathcal{F}$ is Donsker if and only if $\mathcal{F}$ is totally bounded in $\mathcal{L}^2(P)$ and $\mathbb{G}_n$ is asymptotically equicontinuous on $\mathcal{F}$.*

2.3 Vapnik-Červonenkis classes

The requirements shown in the Glivenko-Cantelli and Donsker theorems might be difficult to check for a particular class. Hence, we will study a collection of classes of functions for which the conditions hold: Vapnik-Červonenkis or VC-classes; they are thus called in honor of the two mathematicians who first study them. The importance of this collection lies in the fact that several of the most important classes of functions in statistical applications are VC-classes. Furthermore, for many classes it is relatively straightforward to prove that they belong to this collection.

Definition 2.14. Let $\mathcal{C}$ be a class of subsets of some finite space S. $\mathcal{C}$ is a *VC-class* if there exists a number n such that, from every set of n points in S, the class cannot pick out all of the 2^n distinct subsets (it is said that the class cannot *shatter* the set). Formally, if S consists of n points, then there are strictly less than 2^n distinct sets of the form $S \cap C$ with $C \in \mathcal{C}$. The *VC-index* $V(\mathcal{C})$ of the class is the smallest n for which no set of size n is shattered by $\mathcal{C}$, i.e.

$$V(\mathcal{C}) = \inf\{n : \max_{x_1,...,x_n} \Delta_n(\mathcal{C}, x_1, ..., x_n) < 2^N\},$$

where $\Delta_n(\mathcal{C}, x_1, ..., x_n) = |\{C \cap \{x_1, ..., x_n\} : C \in \mathcal{C}\}|$. It is observed that the VC-index of a class can be ∞, in which case it is not a VC-class.

The main result of this section will give us a bound for the covering number of VC-classes which is good enough to meet the conditions of Donsker theorem. In order to establish it, we will need some previous combinatorial results that will be given without proofs; a detailed account of these facts is given in section 2.6 of van der Vaart and Wellner (1996).

Lemma 2.15. *For a VC-class of sets of index $V(\mathcal{C})$, one has*

$$\max_{x_1,...,x_n} \Delta_n(\mathcal{C}, x_1, ..., x_n) \leq \sum_{j=0}^{V(\mathcal{C})-1} \binom{n}{j}.$$

Consequently, the numbers on the left side grow polynomially of order at most $O(n^{V(\mathcal{C})-1})$ *as* $n \to \infty$.

Theorem 2.16. *There exists a universal constant* K *such that for any VC-class* $\mathcal{C}$ *of sets, any probability measure* Q*, any* $r \geq 1$*, and* $0 < \epsilon < 1$*,*

$$N(\epsilon, \mathcal{C}, L^r(Q)) \leq KV(\mathcal{C})(4e)^{V(\mathcal{C})} \left(\frac{1}{\epsilon}\right)^{r(V(\mathcal{C})-1)}.$$

We want to have a bound such as the one in the previous theorem for classes of functions; looking forward to construct VC-classes of functions, we first define the notion of subgraph of a function.

Definition 2.17. The *subgraph* of a function $f : \mathcal{X} \mapsto \mathbb{R}$ is the subset

$$\{(x,t) : t < f(x)\} \subseteq \mathcal{X} \times \mathbb{R}.$$

Definition 2.18. A class $\mathcal{F}$ of measurable functions is a *VC-subgraph class* if the collection of all subgraphs of the functions in $\mathcal{F}$ is a VC-class of sets. The VC-index of the set of subgraphs of $\mathcal{F}$ is denoted $VC(\mathcal{F})$.

Finally, we can state the main result.

Theorem 2.19. *For a VC-subgraph class with measurable envelope function* F *and* $r \geq 1$*, one has for any probability measure* Q *with* $\|F\|_{Q,r} > 0$*,*

$$N(\epsilon\|F\|_{Q,r}, \mathcal{F}, L^r(Q)) \leq KV(\mathcal{F})(16e)^{V(\mathcal{F})} \left(\frac{1}{\epsilon}\right)^{r(V(\mathcal{C})-1)},$$

for a universal constant K *and* $0 < \epsilon < 1$.

This result shows that for a VC-subgraph class both the condition for covering numbers in the Glivenko-Cantelli theorem and the uniform entropy bound 2.3 hold; this fact implies that a measurable VC-subgraph class $\mathcal{F}$ with envelope function F is $P-$Glivenko-Cantelli for any underlying measure P for which $P^*F < \infty$ and it is $P-$Donsker if $F \in L^2(P)$.

2.4 Examples of VC and Donsker classes

Proving that a class is VC commonly becomes a combinatorial problem; we present some general classes which are VC, and encourage the reader to consult van der Vaart and Wellner (1996) to see the combinatorial methods which are used to check the VC-properties in these cases.

Theorem 2.20. *Any finite-dimensional vector space $\mathcal{F}$ of measurable functions is VC-subgraph with $V(\mathcal{F}) \leq \dim(\mathcal{F}) + 2$.*

Theorem 2.21. *(Parametric families) Let $\mathcal{F} = \{f_\theta : \theta \in \Theta\}$ be a collection of measurable functions indexed by a bounded subset $\Theta \subset \mathbb{R}^d$. Suppose that there exists a measurable function m such that*

$$|f_{\theta_1}(x) - f_{\theta_2}(x)| \leq m(x)\|\theta_1 - \theta_2\|, \text{ for every } \theta_1, \theta_2.$$

If $P|m|^r < \infty$ then $\mathcal{F}$ is a $P-$Donsker class.

The proof of this theorem uses a Donsker theorem defined in terms of the *bracketing entropy*, which is related to the covering numbers. Since this concept and its corresponding Donsker theorem are not necessary for any other result in the present work we will not give more details; we encourage the interested reader to consult Part 2 of van der Vaart and Wellner (1996) and Chapter 19 of van der Vaart (2000).

Now we give a result which give us two important examples of Donsker classes generated from previously given classes.

Theorem 2.22. *Let $\mathcal{F}_1, ..., \mathcal{F}_k$ be uniformly bounded Donsker classes with $\|P\|_{\mathcal{F}_i} < \infty$. If $\phi : \mathbb{R}^k \mapsto \mathbb{R}$ satisfies*

$$|\phi \circ f(x) - \phi \circ g(x)|^2 \leq \sum_{i=1}^{k} (f_i(x) - g_i(x))^2, \tag{2.4}$$

then the class $\phi\circ(\mathcal{F}_1, ..., \mathcal{F}_k)$ is Donsker, provided $\phi\circ(f_1, ..., f_k)$ is square integrable for at least one $(f_1, ..., f_k)$.

Corollary 2.23. *Let $\mathcal{F}$ and $\mathcal{G}$ be uniformly bounded Donsker classes, then the class which consists of the pairwise products $\mathcal{F} \cdot \mathcal{G}$ is Donsker.*

Proof. For $f \in \mathcal{F}$ and $g \in \mathcal{G}$, define $\phi(f,g) = fg$. This function is Lipschitz on bounded subsets of $\mathbb{R}^2$, thus it satisfies 2.4.By applying theorem 2.22 the result is obtained.

Corollary 2.24. *Let $\mathcal{F}$ be a Donsker class with $\|P\|_{\mathcal{F}} < \infty$ and g an uniformly bounded measurable function. Then $\mathcal{F} \cdot g$ is a Donsker class.*

Proof. Once again the function of interest is $\phi(f,g) = fg$. However in this case, we are not considering bounded subsets of $\mathbb{R}^2$, hence ϕ is not Lipschitz on the current domain. Nevertheless, we have

$$|\phi(f_1(x), g(x)) - \phi(f_2(x)g(x))| \leq \|g\|_\infty |f_1(x) - f_2(x)|,$$

for all x. Therefore, condition 2.4 is satisfied and by theorem 2.22 $\mathcal{F} \cdot g$ is a Donsker class.

2.5 Goodness of fit statistics

The specific application of empirical processes theory that will be used in the present work is the testing of goodness of fit, in particular testing for symmetry of a multivariate distribution. In this section we will illustrate the basic process for utilizing the developed theory in developing such tests.

We want to test the null hypothesis that the underlying distribution P of certain data has some kind of symmetry that depends on a parameter θ. In other words we want to test if P belongs to a family $\{P_\theta : \theta \in \Theta\}$ of distributions which have the desired feature. The present approach is to measure the discrepancy between $\mathbb{P}_n$

and $P_{\hat{\theta}}$, where $\hat{\theta}$ is a consistent estimator of θ.

Our statistic for testing symmetry will be a quadratic form of the empirical process and the method in Koltchinskii and Sakhanenko (2000) is based in suprema of it. In general, one can use as statistic functionals of the empirical process, hence the limit distribution will follow by the continuous mapping theorem from the limit distribution of $\sqrt{n}(\mathbb{P}_n - P_{\hat{\theta}})$. The consequence of estimating θ is that the limit process will not be an exact Brownian bridge. The following result gives us the desired limit process.

Definition 2.25. A sequence of estimators $(\hat{\theta}_n)_{n\in\mathbb{N}}$ is *asymptotically linear* with influence function ψ_θ if

$$\sqrt{n}(\hat{\theta}_n - \theta) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi_\theta(X_i) + o_{P_\theta}(1),$$

with $P_\theta\psi_\theta = 0$ and $P_\theta\|\psi_\theta\|^2 < \infty$.

Theorem 2.26. *Let $X_1, ..., X_n$ be a random sample from a distribution P_θ indexed by $\theta \in \mathbb{R}^k$. Let $\mathcal{F}$ be a $P_\theta-$Donsker class of measurable functions and let $(\hat{\theta}_n)_{n\in}$ be a sequence of estimators that are asymptotically linear with influence function ψ_θ. Assume that the map $\theta \mapsto P_\theta$ from $\mathbb{R}^k$ to $l^\infty(\mathcal{F})$ is appropiately differentiable at θ. Then the sequence $\sqrt{n}(\mathbb{P}_n - P_{\hat{\theta}})$ converges under θ in distribution in $l^\infty(\mathcal{F})$ to the process $\mathbb{G}_{P_\theta}f - \mathbb{G}_{P_\theta}\psi_\theta^T\dot{P}_\theta f$.*

Proof. The differentiability of the map $\theta \mapsto P_\theta$ and lemma 2.12 in van der Vaart (2000) imply

$$\|P_{\hat{\theta}_n} - P_\theta - (\hat{\theta} - \theta)^T\dot{P}_\theta\|_{\mathcal{F}} = o_P(\|\theta_n - \theta\|).$$

This let us write

$$n^{1/2}(\mathbb{P}_n - P_{\hat{\theta}}) = n^{1/2}(\mathbb{P}_n - P_\theta) - n^{1/2}(P_{\hat{\theta}} - P_\theta)$$

$$\approx n^{1/2}(\mathbb{P}_n - P_\theta) - n^{1/2}(\hat{\theta} - \theta)\dot{P}_\theta. \tag{2.5}$$

Adding to $\mathcal{F}$ the components of the influence function ψ_{θ} generates a Donsker class, since union of two Donsker classes is Donsker. Thus we can apply the uniform law of large numbers and the Slutsky's lemma to the components of 2.5 to obtain the result.

Chapter 3

A bootstrap method to test for elliptical symmetry and its extension to other symmetries

In this chapter we overview the method developed by Koltchinskii and Sakhanenko in Koltchinskii and Sakhanenko (2000) for testing for elliptical symmetry of a multivariate distribution in which our own method is inspired. Moreover, we discuss the generalization for group symmetry presented by Sakhanenko in Sakhanenko (2009). These methods are nonparametric bootstrap tests whose statistics are defined as functionals of empirical processes indexed by special classes of functions.

3.1 Testing for elliptical symmetry

As stated in Chapter 1, a random vector $\mathbf{X}$ has an *elliptically symmetric* distribution with parameters $\boldsymbol{\theta}_0$ and $\boldsymbol{\Sigma}_0$ if there exists a spherically symmetric random vector $\mathbf{Y}$ such that $\mathbf{X} \stackrel{d}{=} \mathbf{A}_0^T\mathbf{Y} + \boldsymbol{\theta}_0$, where $\mathbf{A}_0$ $k \times d$ satisfies $\mathbf{A}_0^T\mathbf{A}_0 = \boldsymbol{\Sigma}_0$ with $\text{rank}\boldsymbol{\Sigma}_0 = k \leq d$. We will narrow the definition restricting to nonsigular matrices $\mathbf{A}$, in which case $\mathbf{A}^{-1}(\mathbf{X} - \boldsymbol{\theta})$ is spherically symmetric. Although $\mathbf{A}$ is not unique, one can fix it by asking it to be symmetric, positive definite and with norm 1.

Given Π a spherically symmetric Borel probability measure centered about $\mathbf{0}$, a radial measure is defined by

$$\pi_0(\rho) := \Pi(\{x : |x| \leq \rho\}),\ \rho > 0.$$

An elliptically symmetric probability measure P will be identified by the parameters $(\boldsymbol{\theta}_0, \mathbf{A}_0, \pi_0)$ and $E(\mathbb{R}^d)$ will be the set of all elliptically symmetric distributions in $\mathbb{R}^d$.

For a random vector $\mathbf{X}$ with distribution P, not necessarily in $E(\mathbb{R}^d)$, we will suppose that $\mathbb{E}|\mathbf{X}|^2 < \infty$. Then we can define $\boldsymbol{\theta}_0 := \mathbb{E}\mathbf{X} = \int_{\mathbb{R}^d} \mathbf{x} P(d\mathbf{x})$, $\mathbf{A}_0$ as the square root of the covariance operator of $\mathbf{X}$ and π_0 as the distribution of the random variable $\|\mathbf{A}_0^{-1}(\mathbf{X} - \boldsymbol{\theta}_0)\|$. Let us set P^S to be the elliptically symmetric distribution with parameters $(\boldsymbol{\theta}_0, \mathbf{A}_0, \pi_0)$, then P^S will be denoted the elliptical symmetrization of P.

Now we proceed to construct the test statistic. Let m be the uniform distribution on the unit sphere $\mathbb{S}^{d-1}$. For a Borel function f on $\mathbb{R}^d$, we set $m_f(\rho)$ to be an expected value of f relative to the measure m controled by a radial parameter ρ, that is

$$m_f(\rho) := \int_{\mathbb{S}^{d-1}} f(\rho v) m(dv),\ \rho > 0. \tag{3.1}$$

Using the characterization of spherical symmetry, given in Chapter 1, according to which the distributions of $\|\mathbf{X}-\boldsymbol{\theta}\|$ and $\frac{(\mathbf{X}-\boldsymbol{\theta})}{\|\mathbf{X}-\boldsymbol{\theta}\|}$ are independent and the distribution of the latter random vector is uniform over $\mathbb{S}^d$, one can show that for an elliptically symmetric P

$$\int_{\mathbb{R}^d} f(\mathbf{A_0^{-1}}(\mathbf{x} - \theta_\mathbf{0})) P(d\mathbf{x}) = \int_0^\infty m_f(\rho) \pi_0(d\rho). \tag{3.2}$$

Definition 3.1. A class of functions $\mathcal{F}$ *characterizes a distribution* if

$$\int_{\mathbb{R}^d} f dQ_1 = \int_{\mathbb{R}^d} f dQ_2 \text{ for all } f \in \mathcal{F}$$

implies that $Q_1 = Q_2$.

If a class $\mathcal{F}$ characterizes the distribution P, then P is elliptically symmetric iff 3.2 holds for all $f \in \mathcal{F}$. On the contrary, P is $\mathcal{F}-$asymmetrical iff there exists a function $f \in \mathcal{F}$ for which 3.2 does not hold.

Let $(X_1, ..., X_n)$ be a random sample from P, defined in a probability space $(\mathcal{X}, \mathcal{A}, \mathbb{P})$, an let P_n be the empirical distribution based on the given sample. We will use the sample to define the following consistent and $n^{1/2}$-consistent estimators of the parameters $\boldsymbol{\theta}_0$ and $\mathbf{A}_0$ (for consistency of $\mathbf{A}_n$ we need that $\mathbb{E}|X_i|^2 < \infty$, $i = 1, ..., n$):

$$\theta_n := \bar{X}_n := n^{-1} \sum_{i=1}^{n} X_i$$

and

$$\mathbf{A}_n := \left(n^{-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^T (X_i - \bar{X}_n) \right).$$

Definition 3.2. Given a random sample $(X_1, ..., X_n)$, its *scaled residuals* are given by

$$Z_j := \mathbf{A}_n^{-1}(X_j - \theta_n), \; j = 1, ..., n.$$

Now we estimate π_0 with π_n, defined as the empirical distribution based on the sample $(|Z_1|, ..., |Z_n|)$.

The test statistics will be functionals of the stochastic process

$$\xi_n(f) := n^{1/2} \left(\int_{\mathbb{R}^d} f(\mathbf{A}_n^{-1}(\mathbf{x} - \boldsymbol{\theta}_n)) P_n(d\mathbf{x}) - \int_0^\infty m_f(\rho) \pi_n(d\rho) \right) =$$

$$= n^{-1/2} \sum_{j=1}^{n} \left(f(Z_j) - m_f(|Z_j|) \right), \; f \in \mathcal{F},$$

where $\mathcal{F}$ will be a VC-class. Proposed classes $\mathcal{F}$ include indicator functions of the class of cups on the unit sphere, indicator functions of the class of half-spaces and spherical harmonics.

Defining

$$E(f;\boldsymbol{\theta},\mathbf{A}) := \int_{\mathbb{R}^d}[f(\mathbf{A}^{-1}(\mathbf{x}-\boldsymbol{\theta})) - m_f(|\mathbf{A}^{-1}(\mathbf{x}-\boldsymbol{\theta})|)]P(d\mathbf{x})$$

we have by 3.2 that for an elliptically symmetric P and for all f, $E(f,\boldsymbol{\theta}_0,\mathbf{A}_0)=0$. The main theoretical result of Koltchinskii and Sakhanenko (2000) consists in proving the weak convergence of the stochastic process

$$\{\xi_n(f) - n^{1/2}E(f;\boldsymbol{\theta}_0,\mathbf{A}_0):\ f\in\mathcal{F}\}$$

to a Gaussian stochastic process. The formal statement of the result and its proof are omitted since they are quite similar to those that will be presented in Chapter 4. This result implies that $T_n := \|\xi_n\|_{\mathcal{F}}$ has power 1 asymptotically , i.e. if H_0 is the hypothesis that $P\in E(\mathbb{R}^d)$, $H_\alpha(\mathcal{F})$ is the alternative hypothesis that $P\notin E(\mathbb{R}^d)$ and $\mathcal{F}$ characterizes the distribution, then

$\mathbb{P}\{T_n\geq t_\alpha|H_0\}\to\alpha$ and $\mathbb{P}\{T_n\geq t_\alpha|H_\alpha\}\to 1$ as $n\to\infty$.

The limit process might not be known, thus a bootstrap approach is proposed to analyze the value of the statistic. Given the empirical triplet $(\boldsymbol{\theta}_n,\mathbf{A}_n,\pi_n)$, P_n^S denotes the elliptically symmetric distribution with such parameters. A random sample $(X_1^S,...,X_n^S)$ from P_n^S is taken in order to calculate the bootstrapped parameters $\hat{\theta}_n$ and $\hat{\mathbf{A}}_n$. Let $\hat{P}_n$ be the empirical measure based on the sample. The bootstrapped scaled residuals will be

$$\hat{Z}_j := \hat{\mathbf{A}}_n^{-1}(X_j^S-\hat{\theta}_n),\ j=1,...,n.$$

Let $\hat{\pi}_n$ be the empirical distribution based on the sample $(|\hat{Z}_1|,...,|\hat{Z}_n|)$.

The bootstrap version of the process ξ_n is given by

$$\hat{\xi}_n(f) := n^{1/2}\left(\int_{\mathbb{R}^d} f(\hat{\mathbf{A}}_n^{-1}(\mathbf{x}-\hat{\boldsymbol{\theta}}_n))\hat{P}_n(d\mathbf{x}) - \int_0^\infty m_f(\rho)\hat{\pi}_n(d\rho)\right) =$$

$$= n^{-1/2} \sum_{j=1}^{n} \left(f(\hat{Z}_j) - m_f(|\hat{Z}_j|) \right), \; f \in \mathcal{F},$$

The proposal of Koltchinskii and Sakhanenko finishes with the proof of convergence of this bootstrap process, with which the theoretical validity of the method is established.

3.2 Testing for group symmetry

Sakhanenko (2009) proposed an extension of the method presented above to test for group symmetry. Remember from Chapter 1 that a random vector $\mathbf{X}$ has a $\mathcal{S}$-*symmetric* distribution iff there exists an affine nonsigular transformation $\mathbf{A}_0$ such that the random vector $\mathbf{Z} = \mathbf{A}_0\mathbf{X}$ is $\mathcal{S}$-invariant, i.e.

$$\mathbf{Z} \overset{d}{=} \mathbf{A}_0\mathbf{X},$$

for all transformations $S \in \mathcal{S}$. In terms of the measure P this means that there exists a $\mathcal{S}$-invariant Borel probability measure Π_0 such that $P = \Pi_0 \circ \mathbf{A}_0$. Π_0 can fixed to be the distribution of the random variable $\mathbf{Z} = \mathbf{A}_0\mathbf{X}$. $(\mathbf{A}_0, \Pi_0)$ will be called the parameters of the $\mathcal{S}$-symmetric distribution P.

Relation 3.2 is slightly modified in this context; for a $\mathcal{S}$-symmetric distribution with parameters $(\mathbf{A}_0, \Pi_0)$, given any bounded Borel function f we have

$$\int_{\mathbb{R}^d} f(\mathbf{A_0 x}) P(d\mathbf{x}) = \int_{\mathbb{R}^d} m_f(y) \pi_0(dy). \tag{3.3}$$

As in the previous section if $\mathcal{F}$ characterizes the distribution P, then P is $\mathcal{S}$-symmetric iff 3.3 holds for all $f \in \mathcal{F}$. On the contrary, P is $\mathcal{F}-$asymmetrical if there exists a function $f \in \mathcal{F}$ for which 3.3 does not hold. Once again, VC-classes will have the desired properties.

What follows directly corresponds with the steps and definitions of the previous section, then we will not be as detailed. Let $\mathbf{A}_n$ be a $n^{1/2}-$consistent estimator of $\mathbf{A}_0$. The scaled residuals of the observations $(X_1, ..., X_n)$ would be

$$Z_j := \mathbf{A}_n \mathbf{X}_j,\ j = 1, ..., n.$$

The statistics will be sup-norms of the stochastic process

$$\xi_n(f) := n^{1/2} \left(\int_{\mathbb{R}^d} f(\mathbf{A}_n \mathbf{x}) P_n(d\mathbf{x}) - \int_{\mathbb{R}^d} m_f(y) \Pi_n(dy) \right) =$$

$$= n^{-1/2} \sum_{j=1}^{n} \left(f(Z_j) - m_f(Z_j) \right),\ f \in \mathcal{F}.$$

An equivalent convergence theorem is proved for this generalized process, which guarantees the desirable asymptotic power of the test. Moreover, a generalization of the bootstrap process is presented and the corresponding bootstrap process is defined

$$\hat{\xi}_n(f) := n^{1/2} \left(\int_{\mathbb{R}^d} f(\hat{\mathbf{A}}_n \mathbf{x}) \hat{P}_n(d\mathbf{x}) - \int_{\mathbb{R}^d} m_f(y) \hat{\Pi}_n(dy) \right) =$$

$$= n^{-1/2} \sum_{j=1}^{n} \left(f(\hat{Z}_j) - m_f(\hat{Z}_j) \right),\ f \in \mathcal{F}.$$

Where $\hat{P}_n$ is the empirical measure based on the bootstrap sample $(\mathbf{X}_1^S, ..., \mathbf{X}_n^S)$,

$$\hat{\mathbf{A}}_n := \mathbf{A}_n(\mathbf{X}_1^S, ..., \mathbf{X}_n^S)$$

and

$$\hat{\mathbf{Z}}_j := \hat{\mathbf{A}}_n \mathbf{X}_j^S,\ j = 1, ..., n.$$

The convergence of this bootstrap process is also shown, and thus the theoretical foundation for testing for group symmetry is settled.

Chapter 4

Testing for central and angular symmetry

This chapter is concerned with the presentation of the proposed method for testing for central and angular symmetries of a multivariate distribution. Although group symmetry encompasses both, our approach looks forward to use the natural definition of each symmetry for the sake of a more computationally efficient process. The classes of functions and statistics that are used are also proposed with efficiency as a priority.

4.1 Introduction to the method.

We propose a nonparametric method for testing for some symmetries of a multivariate distribution P based on the empirical measure P_n, which is inspired in the ideas presented in Koltchinskii and Sakhanenko (2000). It cannot be used for all group symmetries as the extension presented in Sakhanenko (2009), but the approach is more natural and simple to apply in the cases that are covered. The presentation of this chapter closely follow the layout of Koltchinskii and Sakhanenko (2000), specially in section 4, since the proofs of convergence are quite similar for both methods.

Sakhanenko's generalized method exploits a characterization of symmetry in which the distribution is represented as a product of a group-invariant distribution and an affine nonsingular transformation, i.e. $P = \Pi_0 \circ \mathbf{A_0}$. In contrast, we want to take advantage of the most natural nonparametric definitions of each symmetry. We will develop the method for the case of central symmetry and later we will discuss its logical extension to other forms of symmetry with emphasis in angular symmetry. Let us remember from Chapter 1 that a random vector $\mathbf{X}$ has a centrally symmetric distribution about θ if $\mathbf{X} - \boldsymbol{\theta} \stackrel{d}{=} \boldsymbol{\theta} - \mathbf{X}$.

Following this definition, we see that for a centrally symmetric distribution P, centered in $\boldsymbol{\theta}_0$, and a Borel function f, we have

$$\int_{\mathbb{R}^d} f(\mathbf{x} - \boldsymbol{\theta}_0) P(d\mathbf{x}) = \int_{\mathbb{R}^d} f(\boldsymbol{\theta}_0 - \mathbf{x}) P(d\mathbf{x}) \tag{4.1}$$

If a class of functions $\mathcal{F}$ characterizes the distribution, we have that:

- P is centrally symmetric iff 4.1 holds for all $f \in \mathcal{F}$.

- P is $\mathcal{F}$-asymmetric iff, for some $f \in \mathcal{F}$, 4.1 does not hold.

Since the test is to be nonparametric, we will estimate $\boldsymbol{\theta}_0$ with the mean of the observations

$$\boldsymbol{\theta}_n := \overline{\mathbf{X}}_n := n^{-1} \sum_{i=1}^{n} \mathbf{X}_i$$

which is a consistent and a $n^{1/2}$-consistent estimator.

The test statistics will be quadratic forms of the stochastic process

$$\xi_n(f) := n^{1/2} \left(\int_{\mathbb{R}^d} f(\mathbf{X} - \boldsymbol{\theta}_n) P_n(d\mathbf{x}) - \int_{\mathbb{R}^d} f(\boldsymbol{\theta}_n - \mathbf{X}) P_n(d\mathbf{x}) \right) =$$

$$= n^{-1/2} \sum_{j=1}^{n} \left(f(\mathbf{X}_j - \boldsymbol{\theta}_n) - f(\boldsymbol{\theta}_n - \mathbf{X}_j) \right), \ f \in \mathcal{F},$$

where $\mathcal{F}$ is a class of Borel functions from $\mathbb{R}^d$ into $\mathbb{R}$.

The convergence of this stochastic process will assure the consistency of the statistic and thus, it will prove the validity of the proposed method for testing central symmetry of a multivariate distribution. Nevertheless, before considering the technical aspects, we will further specify the characteristics of the method.

4.2 Testing for central symmetry.

Although we will show that the process ξ_n converges, we will not be able to calculate its limit process, since it may depend on the unknown parameter of P. Consequently, we will evaluate the statistic using bootstrap. We want to construct samples from P_n^S, i.e. the centrally symmetric distribution with parameter θ_n. We take an i.i.d. sample $(\mathbf{X_1}, ..., \mathbf{X_n})$ from P_n, center it by substracting θ_n and multiply each data by -1 or 1 with probability one half for each option; we end the symmetrization process by translating the data to its original center, that is adding $\theta_{\mathbf{n}}$. Summarizing, we have

$$(\mathbf{X}_1^S, ...\mathbf{X}_n^S) := (\epsilon_1(\mathbf{X}_1 - \boldsymbol{\theta}_n), ..., \epsilon_n(\mathbf{X}_n - \boldsymbol{\theta}_n)) + \boldsymbol{\theta}_{\mathbf{n}},$$

where ϵ_i denotes a Rademacher random variable.

By this process we obtain a random sample from P_n^S. Now, let $\hat{P}_n$ be the empirical measure based on this sample and define

$$\hat{\boldsymbol{\theta}}_n := \boldsymbol{\theta}_n(\mathbf{X}_1^S, ...\mathbf{X}_n^S).$$

Finally, we can have a bootstrap version of the process ξ_n

$$\hat{\xi}_n(f) := n^{1/2} \left(\int_{\mathbb{R}^d} f(\mathbf{X} - \hat{\boldsymbol{\theta}}_n) \hat{P}_n^S(d\mathbf{x}) - \int_{\mathbb{R}^d} f(\hat{\boldsymbol{\theta}}_n - \mathbf{X}) \hat{P}_n^S(d\mathbf{x}) \right) =$$

$$= n^{-1/2} \sum_{j=1}^{n} \Big(f(\mathbf{X}_j^S - \hat{\boldsymbol{\theta}}_n) - f(\hat{\boldsymbol{\theta}}_n - \mathbf{X}_j^S) \Big), \ f \in F,$$

We will also prove that this process converges to a limit stochastic process.

Now we discuss the class of functions that will be used for the test. Notice that it suffices to use functions without even parity, since regardless of the underlying distribution, we have $\xi_n = 0$ for every even function f. Therefore, any basis for $\mathcal{F} = \{f : \mathbb{R} \to \mathbb{R} : f(\mathbf{x}) \neq -f(\mathbf{x})\}$ will characterize the distribution of the data. With more information about the underlying distribution of a specific data set we could further reduce the size of the class $\mathcal{F}$, nevertheless we are interested in developing a general method and so we use a class whose applicability does not depend on the data.

Polynomials are a good choice for the current goal, because they are well behaved (they are C^∞, dense in $C[a,b]$ and thus in $L^p,\ p \in [1,\infty)$) and are easy to manipulate and evaluate in a computer. We will use Legendre polynomials because they are an orthonormal basis for polynomials in $L^2[-1,1]$ which consist exclusively of odd and even functions, hence we have to deal only with half of the basis, which is presented up to order 5 in Table 4.1. For data in d dimensions we use polynomials of the form $P_{i,\ldots,j}[x_1, \ldots, x_d] = P_i(x_1) \cdots P_j(x_d)$, where $P_i(x_j)$ is the Legendre polynomial of order i for the variable x_j. That the set of all such products (multivariate Legendre polynomials) is a basis for d dimensional polynomials in $L^2[-1,1]^d$ is derived from the fact that Legendre polynomials are a basis for polynomials in $[-1,1]$.

Table 4.1: Legendre polynomials up to order 5.

$$\begin{array}{rcc}
P_0(x) & = & 1 \\
P_1(x) & = & x \\
P_2(x) & = & \frac{1}{2}(3x^2 - 1) \\
P_3(x) & = & \frac{1}{2}(5x^3 - 3x) \\
P_4(x) & = & \frac{1}{8}(35x^4 - 30x^2 + 3) \\
P_5(x) & = & \frac{1}{8}(63x^5 - 70x^3 + 15x)
\end{array}$$

Although this infinite class of functions is not VC-subgraph, in a real implementation only polynomials up to a certain order can be used, thus we consider a reduced, finite class $\mathcal{F}_k$ of multivariate Legendre polynomials up to order k. Being a finite class of continuous functions, $\mathcal{F}_k$ is VC-subgraph by theorem 2.21, which will allow the theoretical convergence of the process.

An appropiate use of the class $\mathcal{F}_k$ requires the data to be contained into the cube $[-1,1]^d$. For this reason we normalize the data coordinate-wise, that is, we divide the i^{th} coordinate of each point by the maximum norm of that coordinate over all the data points. For heavy-tail distributions this might collapse almost all data into a small vecinity of 0, since the maximum value of a coordinate might come from a point far along a tail. If a problem arises, to avoid this inconvenience, we suggest to trim the data, preserving the points for which all its coordinates belong to the interval between the 5% and the 95% quantiles of the coordinate values. This procedure is forward justify by convergence theorems for quantiles (see González-Barrios and Rueda (2001)), which guarantee the convergence of these intervals to a limit value. Nevertheless, this process has to be done carefully and only if it is of upmost necessity, because the trimming of the data may symmetrize an otherwise asymmetric sample.

Finally, we establish the statistic that will be used for testing the symmetry hypothesis. Although we could implement a Kolmogorov-Smirnov type test, for which the asymptotic power is 1 as proved in Koltchinskii and Sakhanenko (2000), we decide to use a statistic based on a quadratic form of the process ξ_n because this kind of statistics has shown a better behaviour in implementations for small data sets (for example see Manzotti and Quiroz (2001)).

Let n be the number of data points and k be the number of functions in the class $\mathcal{F}_k$ which will be evaluated. Denote by ν_n the vector of length k whose component i corresponds to the evaluation of the process ξ_n through all the data for the i^{th} function of $\mathcal{F}$. Let m be the number of bootstrap samples that will be use and construct for each sample j the vector $\nu_{n,j}$, i.e. the equivalent of vector ν_n for the bootstrap sample, where the points are used to evaluate the bootstrap process $\hat{\xi}_n$ instead of the original process.

Let $\hat{\mathbf{C}}$ be the covariance matrix of the m vectors ν_j defined previously. Our statistic Q_n will be

$$Q_n := \nu_n^t \hat{\mathbf{C}} \nu_n.$$

In order to test the null hypothesis of central symmetry H_0 of the distribution at level α we will calculate the value $Q_{n,j} = \nu_{n,j}^t \hat{\boldsymbol{C}} \nu_{n,j}$ for each bootstrap sample and take the $1-\alpha$ quantile $q_{1-\alpha}$ from this m data. If $Q_n \leq q_{1-\alpha}$ the alternative hypothesis will be rejected at confidence level α and if $Q_n > q_{1-\alpha}$ it will not be rejected. To calculate the power of the test against a specific alternative H_α, we will repeat the process l times counting the number s of cases in which H_α is accepted. From this, the power of the test against H_α is given by s/l.

4.3 Proofs

We first define the required convergence as stated in Koltchinskii and Sakhanenko (2000).

Definition 4.1. A sequence of stochastic processes $\zeta_n : \mathcal{F} \to \mathbb{R}$ *converges weakly* in $l^\infty(\mathcal{F})$ to a stochastic process $\zeta : \mathcal{F} \to \mathbb{R}$ iff there exists a Radon probability measure γ on $l^\infty(\mathcal{F})$ such that γ is the distribution of ζ and, for all bounded and $\|\cdot\|_F$-continuous functionals $\Phi : l^\infty(\mathcal{F}) \to \mathbb{R}$, we have $\mathbb{E}^*\Phi(\zeta_n) \to \int \Phi(x)\gamma(dx)$.

Let us state the two main theorems upon which the validity of the method rests. The first deal with the convergence of the stochastic process from which the statistic is calculated, whereas the second extends this result to the bootstrap process.

Theorem 4.2. *Let $\mathcal{F}$ be a uniformly bounded VC-subgraph class, suppose that smoothness conditions (S) hold and $\int_{\mathbb{R}^d} |x|^4 dP < \infty$. Then the sequence of stochastic processes*

$$\left\{\xi_n(f) - n^{1/2} E(f, \boldsymbol{\theta}_0) : f \in \mathcal{F}\right\}$$

converges weakly in the space $l^\infty(\mathcal{F})$ to a Gaussian stochastic process ξ_P. In particular, if P is centrally symmetric with parameter θ_0, then the sequence $\{\xi_n\}$ converges weakly in the space $l^\infty(\mathcal{F})$ to the process ξ_P.

Before stating the second theorem we give some definitions. Consider the set of all functionals $\Phi : l^\infty(\mathcal{F}) \to \mathbb{R}$ such that for all $Y \in l^\infty(F)$, $|\Phi(Y)| \leq 1$ and for all $Y_1, Y_2 \in l^\infty(\mathcal{F})$, $|\Phi(Y_1) - \Phi(Y_2)| \leq \| Y_1 - Y_2 \|_F$. Call this set $BL_1(l^\infty(\mathcal{F}))$. For two stochastic processes $\zeta_1, \zeta_2 : \Sigma \times \hat{\Sigma} \times \mathcal{F} \to \mathbb{R}$ the following bounded Lipschitz distance is defined

$$d_{BL}(\zeta_1, \zeta_2) := \sup_{\Phi \in BL_1(l^\infty(F))} |\hat{E}^* \Phi(\zeta_1) - \hat{E}^* \Phi(\zeta_2)|.$$

Definition 4.3. A sequence of stochastic processes $\{\xi_n\}$ converges weakly in the space $l^\infty(\mathcal{F})$ to a version $\hat{\xi}_{PS}$ of the process ξ_{PS} in probability $\mathbb{P}$ iff

$$d_B L(\hat{\xi}_n; \hat{\xi}_{PS}) \longrightarrow 0 \text{ as } n \to \infty \text{ in probability } \mathbb{P}.$$

Theorem 4.4. *Let F be a uniformly bounded VC-subgraph class, suppose that smoothness conditions (S) hold and $\int_{\mathbb{R}^d} |x|^4 dP < \infty$. Then the sequence of stochastic processes $\{\hat{\xi}_n\}$ converges weakly in the space $l^\infty(\mathcal{F})$ to a version $\hat{\xi}_{PS}$ of the process ξ_{PS} in probability $\mathbb{P}$. In particular, if P is centrally symmetric with parameter θ_0, then the sequence $\{\hat{\xi}_n\}$ converges weakly to a version of the process ξ_P.*

Before proving these results, let us present some notational abbreviation and a preliminary lemma. For a function f on $\mathbb{R}^d$, define

$$\tilde{f}_{\boldsymbol{\theta}}(x) = f(\mathbf{x} - \boldsymbol{\theta}) - f(\boldsymbol{\theta} - \mathbf{x}).$$

For a class of functions $\mathcal{F}$, define

$$\tilde{\mathcal{F}} := \{\tilde{f}_{\boldsymbol{\theta}} : f \in \mathcal{F}, \boldsymbol{\theta} \in \mathbb{R}^d\}$$

With this abbreviation the processes ξ_n and $\hat{\xi}_n$ are respectively given by

$$\xi_n(f) = n^{1/2} \int_{\mathbb{R}^d} \tilde{f}_{\boldsymbol{\theta}_n}(\mathbf{x}) dP_n, \ \ f \in \mathcal{F}$$

and

$$\hat{\xi}_n(f) = n^{1/2} \int_{\mathbb{R}^d} \tilde{f}_{\hat{\boldsymbol{\theta}}_n(\mathbf{x})} d\hat{P}_n, \ \ f \in \mathcal{F}.$$

Furthermore,

$$E(f; \boldsymbol{\theta}) = \int_{\mathbb{R}^d} \tilde{f}_{\boldsymbol{\theta}} dP$$

For the samples of the symmetric distribution with parameter $\boldsymbol{\theta}$, we define

$$E^S(f; \boldsymbol{\theta}) = \int_{\mathbb{R}^d} \tilde{f}_{\boldsymbol{\theta}} dP^S$$

For the symmetrized measure we write, given a function g on $\mathbb{R}^d$,

$$\int_{\mathbb{R}} g(\mathbf{x}) P_n^S(dx) = \int_{\mathbb{R}} \int_{\{-1,1\}} g(\epsilon \mathbf{x} + \boldsymbol{\theta}_n) R(d\epsilon) P_n(d\mathbf{x}),$$

where $R(d\epsilon)$ is the measure related to a Rademacher random variable ϵ i.e. it assigns to ϵ a 1 or a -1 with probability 0.5 for each possibility. We can write

$$\int_{\mathbb{R}} g(\mathbf{x}) P_n^S(d\mathbf{x}) = \int_{\mathbb{R}} M_g(\boldsymbol{\theta}_n; \mathbf{x}) P_n(d\mathbf{x}),$$

with $M_g(\boldsymbol{\theta}_n; \mathbf{x}) := \int_{\{-1,1\}} g(\epsilon \mathbf{x} + \boldsymbol{\theta}_n) R(d\epsilon)$.

Equivalently, we obtain

$$\int_{\mathbb{R}} g(\mathbf{x}) P^S(d\mathbf{x}) = \int_{\mathbb{R}} M_g(\boldsymbol{\theta}_0; \mathbf{x}) P(d\mathbf{x}).$$

Define

$$\Gamma(g; \boldsymbol{\theta}) := \int_{\mathbb{R}} M_g(\boldsymbol{\theta}_0; \mathbf{x}) P(d\mathbf{x}).$$

For $\mathcal{G}$ a class of functions, denote

$$\mathcal{G}^2 := \{gh : g, h \in \mathcal{G}\},$$

$$M(\mathcal{G}) := \{M_g(\boldsymbol{\theta}, \cdot) : g \in \mathbf{G}, \boldsymbol{\theta} \in \mathbb{R}^d\}.$$

We will need the following lemmas from functional analysis.

Lemma 4.5. *If P and $\mathcal{F}$ satisfy the smoothness conditions (S), the following statements hold:*

1. *If $\boldsymbol{\theta} \longrightarrow \boldsymbol{\theta}_0$ then,*

$$\sup_{f \in \mathcal{F}} \int_{\mathbb{R}^d} |\tilde{f}_{\boldsymbol{\theta}} - \tilde{f}_{\boldsymbol{\theta}_0}|^2 dP \longrightarrow 0$$

   and

$$\sup_{f \in \mathcal{F}} \int_{\mathbb{R}^d} |\tilde{f}_{\boldsymbol{\theta}} - \tilde{f}_{\boldsymbol{\theta}_0}|^2 dP^S \longrightarrow 0$$

2. *The function $E(f; \boldsymbol{\theta})$ is differentiable at the point $(\boldsymbol{\theta}_0)$ for any $f \in \mathcal{F}$, and the Taylor expansion of the first order*

$$E(f; \boldsymbol{\theta}) = E(f; \boldsymbol{\theta}_0) + E'_{\boldsymbol{\theta}}(f; \boldsymbol{\theta}_0)(\boldsymbol{\theta} - \boldsymbol{\theta}_0) + o(|\boldsymbol{\theta} - \boldsymbol{\theta}_0|)$$

   holds uniformly in $f \in \mathcal{F}$.

3. *The function $E^S(f; \boldsymbol{\theta})$ is differentiable at the point $(\boldsymbol{\theta}_0)$ for any $f \in F$, and the Taylor expansion of the first order*

$$E^S(f;\boldsymbol{\theta}) = E^S(f;\boldsymbol{\theta}_0) + (E^S)'_{\boldsymbol{\theta}}(f;\boldsymbol{\theta}_0)(\boldsymbol{\theta}-\boldsymbol{\theta}_0) + o(|\boldsymbol{\theta}-\boldsymbol{\theta}_0|)$$

holds uniformly in $f \in \mathcal{F}$.

4. *The function $\Gamma(g;\boldsymbol{\theta})$ is continuous with respect to $\boldsymbol{\theta}$ at $\boldsymbol{\theta}_0$ uniformly in $g \in (\tilde{\mathcal{F}})^2$.*

5. *The function $\Gamma(g;\boldsymbol{\theta})$ is differentiable at the point $\boldsymbol{\theta}_0$ for any $g \in \tilde{\mathcal{F}}$ and the Taylor expansion of first order*

$$\Gamma(g;\boldsymbol{\theta}) = \Gamma(g;\boldsymbol{\theta}_0) + \Gamma'_{\boldsymbol{\theta}}(g;\boldsymbol{\theta}_0)(\boldsymbol{\theta}-\boldsymbol{\theta}_0) + o(|\boldsymbol{\theta}-\boldsymbol{\theta}_0|)$$

holds uniformly in $g \in \tilde{\mathcal{F}}$.

6. *If $\boldsymbol{\theta} \to \boldsymbol{\theta}_0$, then*

$$\sup_{g\in\tilde{\mathcal{F}}} \int_{\mathbb{R}^d} |M_g(\boldsymbol{\theta};\mathbf{x}) - M_g(\boldsymbol{\theta}_0;\mathbf{x})|^2 P(d\mathbf{x}) \to 0$$

and

$$\sup_{f\in\mathcal{F}} \int_{\mathbb{R}^d} |M_{(\tilde{f})_{\boldsymbol{\theta}}}(\boldsymbol{\theta};\mathbf{x}) - M_{(\tilde{f})_{\boldsymbol{\theta}_0}}(\boldsymbol{\theta}_0;\mathbf{x})|^2 P(d\mathbf{x}) \to 0$$

We will prove convergence properties for the classes of functions defined based on $\mathcal{F}$.

Lemma 4.6. *The classes of function $\tilde{\mathcal{F}}$ and $M(\tilde{\mathcal{F}})$ are uniformly Donsker. Moreover, the class $M(\tilde{\mathcal{F}}^2)$ is uniformly Glivenko-Cantelli.*

Proof. We have that $\mathcal{F}$ is VC-subgraph. By theorem 2.22 for parametric families, $\tilde{\mathcal{F}}$ is VC-subgraph. Considering the definition of $M_g(\boldsymbol{\theta}_n;\mathbf{x})$, the class $M(\tilde{\mathcal{F}})$

consists of integrals over a Radamacher measure of functions translated by a parameter. Integrating accounts for multiplying the function by 1 or -1, thus the size of the class is of the same order of that of the family of functions of the form $\tilde{f}(\epsilon\mathbf{x} + \boldsymbol{\theta}_n)$. Since $\boldsymbol{\theta}_n$ belongs to a parametric family, once again by theorem 2.22 this class is VC-subgraph, hence $M(\tilde{\mathcal{F}})$ is VC-subgraph as well; corollary 2.25 could have been use to prove this if one interprets the integrals as a product by a bounded function. Finally, by corollary 2.24, $\tilde{\mathcal{F}}^2$ is VC-subgraph and by the same consideration made previously, $M(\tilde{\mathcal{F}}^2)$ is VC-subgraph. Since the three classes of interest are VC-subgraph, we obtain the results.

Proof of Theorem 4.2

We define a new process

$$\eta_n(f;\boldsymbol{\theta}) := n^{1/2}(P_n - P)(\tilde{f}_{\boldsymbol{\theta}}),\ f \in \mathcal{F},\ \boldsymbol{\theta} \in \mathbb{R}^d.$$

$\tilde{\mathcal{F}}$ being a P-Donsker class, and taking into account statement 1 of lemma 4.5, by asymptotic equicontinuity we obtain for all ϵ

$$\lim_{\delta\to 0}\limsup_{n\to\infty} \mathbb{P}^*(\{\sup_{|\boldsymbol{\theta}-\boldsymbol{\theta}_0|\le\delta}\ \sup_{f\in F} |\eta_n(f;\boldsymbol{\theta}) - \eta_n(f;\boldsymbol{\theta}_0)| \ge \epsilon\}) = 0. \tag{4.2}$$

Let us now consider the process of our interest, for which we have

$$\xi_n(f) - n^{1/2}E(f;\boldsymbol{\theta}_0) = \eta_n(f;\boldsymbol{\theta}_n) + n^{1/2}(E(f;\boldsymbol{\theta}_n) - E(f;\boldsymbol{\theta}_0)) =$$

$$\eta_n(f;\boldsymbol{\theta}_0) + n^{1/2}(E(f;\boldsymbol{\theta}_n) - E(f;\boldsymbol{\theta}_0)) + (\eta_n(f;\boldsymbol{\theta}_n) - \eta_n(f;\boldsymbol{\theta}_0)).$$

We can use this last addend to bound the suprema of the other part of the expression; if $|\boldsymbol{\theta}_n - \boldsymbol{\theta}_0| \le \delta$, we obtain

$$\sup_{f\in F} |\xi_n(f) - n^{1/2}E(f;\boldsymbol{\theta}_0) - \eta_n(f;\boldsymbol{\theta}_0) -$$

$$-n^{1/2}(E(f;\boldsymbol{\theta}_n) - E(f;\boldsymbol{\theta}_0))| \le$$

$$\leq \sup_{|\boldsymbol{\theta}-\boldsymbol{\theta}_0|\leq\delta} \sup_{f\in F} |\eta_n(f;\boldsymbol{\theta}) - \eta_n(f;\boldsymbol{\theta}_0)|.$$

From the last equation, 4.2 and consistency of the statistic θ_n, we arrive to (uniformly in $f \in \mathcal{F}$)

$$\xi_n(f) - n^{1/2}E(f;\boldsymbol{\theta}_0) =$$

$$= n^{1/2}\int_{\mathbb{R}^d} \tilde{f}_{\boldsymbol{\theta}_0} d(P_n - P) + n^{1/2}(E(f;\boldsymbol{\theta}_n) - E(f;\boldsymbol{\theta}_0)) + o_p(1). \tag{4.3}$$

We have used the notation

$$\eta_n(f;\boldsymbol{\theta}_0) = n^{1/2}\int_{\mathbb{R}^d} \tilde{f}_{\boldsymbol{\theta}_0} d(P_n - P).$$

From the Taylor represention given in statement 2 of lemma 4.5, and due to $n^{1/2}$-consistency of $\boldsymbol{\theta}_n$, we have that uniformly in $f \in \mathcal{F}$

$$n^{1/2}(E(f;\boldsymbol{\theta}_n) - E(f;\boldsymbol{\theta}_0)) =$$

$$= n^{1/2}E'_{\boldsymbol{\theta}}(f,\boldsymbol{\theta}_0)(\boldsymbol{\theta}_n - \boldsymbol{\theta}) + o_p(1),\ n \longrightarrow \infty. \tag{4.4}$$

Replacing 4.4 in 4.3 and using statement 1 of lemma 4.5, we obtain the weak convergence of $\xi_n(f) - n^{1/2}E(f;\boldsymbol{\theta}_0)$ in $l^\infty(\mathcal{F})$ to the Gaussian stochastic process given by

$$\xi_P(f) = \mathbb{G}_P\tilde{f}_{\boldsymbol{\theta}_0} + E'_{\boldsymbol{\theta}}(f,\boldsymbol{\theta}_0)\mathbb{G}_P(x),\ f \in F,$$

where $\mathbb{G}_P$ is the P-Brownian bridge defined in Chapter 2.

When P is centrally symmetric $E(f;\boldsymbol{\theta}_0) = 0$ and we have the convergence of ξ_n to $\xi_P = \mathbb{G}_P\tilde{f}_{\boldsymbol{\theta}_0}$.

Proof of Theorem 4.4. Consider the process

$$\hat{\eta}_n(f;\boldsymbol{\theta}) := n^{1/2}(\hat{P}_n - P_n^S)(\tilde{f}_{\boldsymbol{\theta}}),\ f \in F,\ \boldsymbol{\theta} \in \mathbb{R}^d$$

Statement 4 of lemma 4.5 and the fact that $M((\tilde{\mathcal{F}})^2)$ is uniformly Glivenko-Cantelli imply that uniformly in $g, h \in \tilde{F}$

$$\int_{\mathbb{R}^d} ghdP_n^S \to \int_{\mathbb{R}^d} ghdP^S \text{ as } n \to \infty \text{ a.s}$$

Equivalently for $M(\tilde{\mathcal{F}})$, by statement 5 of lemma 4.5

$$\int_{\mathbb{R}^d} gdP_n^S \to \int_{\mathbb{R}^d} gdP^S \text{ as } n \to \infty \text{ a.s}$$

Then, since $\tilde{\mathcal{F}}$ is uniform Donsker, a.s $n^{1/2}(\hat{P}_n - P_n^S)$ and $n^{1/2}(\tilde{P}_n - P^S)$ converge weakly in the space $l^\infty(\tilde{F})$ to the same limit, the P^S-Brownian bridge $\mathbb{G}_{P^S}$ (This follows from Corollary 2.7 in Giné and Zinn (1991)). Using statement 1 of Lemma 4.5 and asymptotic equicontinuity yields for all $\epsilon > 0$ $\mathbb{P}$ a.s.

$$\lim_{\delta \to 0} \limsup_{n \to \infty} \hat{\mathbb{P}}^*(\{\sup_{|\boldsymbol{\theta} - \boldsymbol{\theta}_n| \leq \delta} \sup_{f \in F} |\hat{\eta}_n(f; \boldsymbol{\theta}) - \hat{\eta}_n(f; \boldsymbol{\theta}_n)| \geq \epsilon\}) = 0. \tag{4.5}$$

Let

$$\hat{E}_n(f; \boldsymbol{\theta}) := \int_{\mathbb{R}^d} \tilde{f}_{\boldsymbol{\theta}} dP_n^S.$$

By construction $\hat{E}_n(f; \boldsymbol{\theta}_n) = 0$, thus we have

$$\hat{\xi}_n(f) = \hat{\eta}_n(f; \hat{\boldsymbol{\theta}}_n) + n^{1/2}(\hat{E}_n(f; \hat{\boldsymbol{\theta}}_n) =$$

$$\hat{\eta}_n(f; \boldsymbol{\theta}_n) + n^{1/2}(\hat{E}_n(f; \hat{\boldsymbol{\theta}}_n) - \hat{E}_n(f; \boldsymbol{\theta}_n)) + (\hat{\eta}_n(f; \hat{\boldsymbol{\theta}}_n) - \hat{\eta}_n(f; \boldsymbol{\theta}_n)).$$

Taking suprema, when $|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_n| \leq \delta$ we obtain

$$\sup_{f \in F} |\hat{\xi}_n(f) - \hat{\eta}_n(f; \boldsymbol{\theta}_n) -$$

$$-n^{1/2}(\hat{E}_n(f; \hat{\boldsymbol{\theta}}_n) - \hat{E}_n(f; \boldsymbol{\theta}_n))| \leq$$

$$\leq \sup_{|\boldsymbol{\theta}-\boldsymbol{\theta}_0|\leq\delta} \sup_{f\in F} |\hat{\eta}_n(f;\boldsymbol{\theta}) - \hat{\eta}_n(f;\boldsymbol{\theta}_n)|. \tag{4.6}$$

From 4.6, 4.5 and consistency of the statistics $\boldsymbol{\theta}_n$ and $\hat{\boldsymbol{\theta}}_n$, we arrive to (uniformly in $f \in F$)

$$\hat{\xi}_n(f) = n^{1/2} \int_{\mathbb{R}^d} \tilde{f}_{\boldsymbol{\theta}_n} d(\hat{P}_n - P_n^S) + n^{1/2}(\hat{E}_n(f;\hat{\boldsymbol{\theta}}_n) - \hat{E}_n(f;\boldsymbol{\theta}_n)) + o_p(1). \tag{4.7}$$

Once again, asymptotic equicontinuity for $n^{1/2}(\hat{P}_n - P_n^S)$ and statement 1 of Lemma 4.5 yield for $n \longrightarrow \infty$

$$n^{1/2} \int_{\mathbb{R}^d} \tilde{f}_{\boldsymbol{\theta}_n} d(\hat{P}_n - P_n^S) = n^{1/2} \int_{\mathbb{R}^d} (\tilde{f})_{\boldsymbol{\theta}_0} d(\hat{P}_n - P_n^S) + o_P(1) \tag{4.8}$$

Now we focus on the second term on the right of 4.7

$$n^{1/2}(\hat{E}_n(f;\hat{\boldsymbol{\theta}}_n) - \hat{E}_n(f;\boldsymbol{\theta}_n)) = n^{1/2}(\hat{E}_n(f;\hat{\boldsymbol{\theta}}_n) - E^S(f;\hat{\boldsymbol{\theta}}_n)) -$$

$$- n^{1/2}(\hat{E}_n(f;\boldsymbol{\theta}_n) - E^S(f;\boldsymbol{\theta}_n)) + n^{1/2}(E^S(f;\hat{\boldsymbol{\theta}}_n) - E^S(f;\boldsymbol{\theta}_n)). \tag{4.9}$$

For the first and second terms of this expression we have

$$n^{1/2}(\hat{E}(f;\hat{\boldsymbol{\theta}}) - E^S(f;\hat{\boldsymbol{\theta}})) =$$

$$n^{1/2}(P_n - P)(M_{\tilde{f}_{\boldsymbol{\theta}}}(\boldsymbol{\theta}_n)) + n^{1/2}(\Gamma((\tilde{f})_{\boldsymbol{\theta}};\boldsymbol{\theta}_n) - \Gamma(\tilde{f}_{\boldsymbol{\theta}};\boldsymbol{\theta}_0)) =$$

$$= n^{1/2}(P_n - P)(M_{\tilde{f}_{\boldsymbol{\theta}}}(\boldsymbol{\theta}_0)) +$$

$$[n^{1/2}(P_n - P)(M_{\tilde{f}_{\boldsymbol{\theta}}}(\boldsymbol{\theta}_n)) - n^{1/2}(P_n - P)(M_{\tilde{f}_{\boldsymbol{\theta}}}(\boldsymbol{\theta}_0))] +$$

$$+ n^{1/2}(\Gamma(\tilde{f}_{\boldsymbol{\theta}};\boldsymbol{\theta}_n) - \Gamma(\tilde{f}_{\boldsymbol{\theta}};\boldsymbol{\theta}_0)).$$

From this representation, by statements 5 and 6 of lemma 4.5 and considering that $M(\tilde{\mathcal{F}})$ is uniformly Donsker, it can be shown that the process $n^{1/2}(\hat{E}_n(f;\boldsymbol{\theta}_n) - E^S(f;\boldsymbol{\theta}_n))$, for $\boldsymbol{\theta} \in \mathbb{R}^d$ weakly converges in the space $l^\infty(\mathcal{F} \times \mathbb{R}^d))$. This, in conjunction with asymptotic equicontinuity and statement 6 of lemma 4.5, give that as $n \longrightarrow \infty$

$$n^{1/2}(\hat{E}_n(f;\hat{\boldsymbol{\theta}}_n) - E_n^S(f;\hat{\boldsymbol{\theta}}_n)) -$$

$$- n^{1/2}(\hat{E}_n(f;\boldsymbol{\theta}_n) - E^S(f;\boldsymbol{\theta}_n)) = o_p(1). \tag{4.10}$$

From the Taylor representation for $E_n^S(f;\hat{\boldsymbol{\theta}}_n)$ given in statement 3 of lemma 4.5, and taking into account the $n^{1/2}$-consistency of $\boldsymbol{\theta}_n$ and $\hat{\boldsymbol{\theta}}_n$, we get uniformly in $f \in F$ as $n \longrightarrow \infty$

$$n^{1/2}(E_n^S(f;\hat{\boldsymbol{\theta}}_n) - E^S(f;\boldsymbol{\theta}_n)) =$$

$$= n^{1/2}(E^S)'_{\boldsymbol{\theta}}(f,\boldsymbol{\theta}_0)(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_n) + o_p(1) =$$

$$= (E^S)'_{\boldsymbol{\theta}}(f,\boldsymbol{\theta}_0)(n^{1/2}\int_{\mathbb{R}^d} xd(\hat{P}_n - P_n^S)) + o_p(1). \tag{4.11}$$

From equations 4.7-4.11 and again using Corollary 2.7 in Giné and Zinn (1991) we get the desired convergence.

Using the Continuous Mapping Theorem we obtain the convergence of the quadratic forms based on ξ_n and $\hat{\xi}_n$ to squares of the Brownian bridges found above. Nevertheless, the method is not based in the exact process to which convergence is achieved, since it can depend in the unknown parameters of the distribution being tested. For this reason we use bootstrap samples when evaluating the statistic.

4.4 Testing for angular symmetry

The method proposed for testing for central symmetry of a multivariate distribution can be adapted for other kinds of symmetries. We will detail the extension of the method for angular symmetry and give the general idea for other possibilities. Remember from Chapter 1 that a distribution has *angular symmetry* iff $\frac{\mathbf{X}-\boldsymbol{\theta}}{\|\mathbf{X}-\boldsymbol{\theta}\|} \stackrel{d}{=} \frac{\boldsymbol{\theta}-\mathbf{X}}{\|\mathbf{X}-\boldsymbol{\theta}\|}$.

Notice from the definition that angular symmetry can be thought as a normalized central symmetry, and that all steps in the method can be implemented with only small changes. In particular, the process in which the statistic will be based is

$$\xi_n(f) := n^{1/2}\left(\int_{\mathbb{R}^d} f(\frac{\mathbf{X}-\boldsymbol{\theta}_n}{\|\mathbf{X}-\boldsymbol{\theta}\|})P_n(dx) - \int_{\mathbb{R}^d} f(\frac{\boldsymbol{\theta}_n-\mathbf{X}}{\|\mathbf{X}-\boldsymbol{\theta}\|})P_n(dx)\right) =$$

$$= n^{-1/2}\sum_{j=1}^{n}\left(f(\frac{\mathbf{X}_j-\boldsymbol{\theta}_n}{\|\mathbf{X}-\boldsymbol{\theta}\|}) - f(\frac{\boldsymbol{\theta}_n-\mathbf{X}_j}{\|\mathbf{X}-\boldsymbol{\theta}\|})\right),\ f \in F.$$

The proofs for convergence are almost identical, as well as the details of the statistic and the implementation, thus we will not delve deeper in these topics. The construction of symmetrized samples through bootstrap is exactly the same, but taking into account the projection of the data onto the unit sphere Σ^d. Even the class F_k of multivariate Legendre polynomials up to degree k is appropiate for this test, since the angular symmetry is also characterized by its even parity. However, we propose another class of functions which might work better for this specific symmetry.

We will use the the class H_k of spherical harmonics up to degree k, where a spherical harmonic of degree j is the restriction to the unitary sphere Σ^d of a homogeneous polynomial $p(x)$ of degree j on $\mathbb{R}^d$, such that $\Delta(p) \equiv 0$ on $\mathbb{R}^d$. Formulae for spherical harmonics up to order 4 in $\mathbb{R}^d$ and a detailed account of its derivation is found in Manzotti and Quiroz (2001).

The selection of spherical harmonics for testing for angular symmetry is mainly based in three reasons:

- They are a well-behaved orthonormal basis for a set which is dense in the continuous functions over the unitary sphere Σ^d, such as the Legendre polynomials were for functions in $[-1,1]^d$.

- For a chosen degree k, the number of spherical harmonics that have to be evaluated are less than the number of multivariate Legendre polynomials, and it grows slower with increasing dimension. The number of linearly independent spherical harmonics of degree j in dimension d is

$$SH(d,j) = \binom{d+j-1}{j} - \binom{d+j-3}{j-2},$$

  whereas the correspondent number of multivariate Legendre polynomials is

$$LP(d,j) = \binom{d+j-1}{d-1}$$

- They have shown good performance when used for similar statistical tests, e.g. see Quiroz and Dudley and Manzotti and Quiroz (2001).

4.5 Testing for other symmetries.

The developed method can be extended to other symmetries under some considerations. The general idea of the method is to select a symmetry which can be defined by $\mathbf{X} \stackrel{d}{=} T\mathbf{X}$, where T is an appropriate transformation which defines the symmetry (it can be seen as the action of a group over $\mathbf{X}$); and evaluate a statistic based on the process

$$\xi_n(f) := n^{1/2}\left(\int_{\mathbb{R}^d} f(X)P_n(dx) - \int_{\mathbb{R}^d} f(TX)P_n(dx)\right),\ f \in F.$$

For central and angular symmetries, the transformation consists in multiplying the given variable by -1, which allows the direct evaluation of the functions in the transformed data. However, the transformation might be more complex, e.g. a distribution has sign symmetry iff $(X_1-\theta_1, ..., X_d-\theta_d) \stackrel{d}{=} (\pm(X_1-\theta_1), ..., \pm(X_d-\theta_d))$, for every combination of $+$ and $-$. In this case, the functions cannot be readily evaluated since we have 2^d possibilities for the point in the second integral. The method can be used to test independently for each combination of signs (central symmetry corresponds to the combination where all signs are $-$) and the null hypothesis of sign symmetry would be rejected if anyone of the sign combinations is rejected. Clearly this is not an efficient method, although it has the advantage that it can discover particular symmetries for some sign combinations even if the distribution does not have sign symmetry. This process can be used to test for any symmetry in which the transfomation applied to the random variable $\mathbf{X}$ consists of a finite number of possibilities; in other words, when the group which determines the symmetry is finite, e.g. thge group of symmetry determined by a polygon.

Another possibility for the case of sign symmetry is to slightly modify the process by uniformly selecting a combination of signs for every point in which the process is evaluated; once a combination is selected the second integral in the process can be calculated. That would affect the rate of convergence of the process because for each point only one of many combinations is being taken into account; but, since the symmetrized samples would be constructed in an equivalent fashion, the method would still be able to distinguish the symmetry features of the distribution. Once again, this idea can be extended to any symmetry whose group representation is finite. Nevertheless, we do not further investigate these modifications to the original method, since they would have particular features for each symmetry which would have to be exploited in order to develop an efficient test. General methods for testing for group symmetries can be found in Sakhanenko (2009).

Chapter 5

Results and Analysis

The tests for central and angular symmetry were implemented in R. We consideral several samples taken from symmetric and asymmetric distributions in order to verify that the method can distinguish the symmetry properties of given data.

5.1 Central symmetry

For central symmetry the method was implemented with Legendre polynomials up to order 3. For large dimensions the number of such polynomials is too high, thus the computational efficiency of the program is strongly reduced. However, since only polynomials of odd degree are of interest, as explained in Chapter 4, we can only reduce the number of polynomials by exclusively considering those of order one, but that would imply a great loss of information about the sample. For bivariate samples we have six different functions to be evaluated. In dimension five ($d = 5$) this number grows to 40. Since the slowest process of the program is the evaluation of the functions, we had to reduce the number of simulations when considering five dimensional samples.

For $d = 2$ we considered $1,000$ bootstrap samples with sizes $n = 100$ and 200. The test was performed with a 5% significance level. We calculated the statistic 100 times for each sample and used the obtained quantiles to find the simulation level for symmetric distributions and the power against asymmetric alternatives. For $d = 5$ we reduced the number of bootstrap samples to 200 with $n = 50$; although this could seem to be a low number of bootstrap samples, affecting the precision of the statistic, the method performed well as is explained below.

We consider four different symmetric distributions to validate the method; when only a univariate distribution is given, it corresponds to the distribution of the marginals.

- S_1: $Unif(-1, 1)$.

- S_2: $\beta(2, 2)$.

- S_3: $\mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_0)$, bivariate normal distribution with mean $\mathbf{0}$ and covariance matrix $\mathbf{\Sigma}_0 = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$.

- S_4: A 50-50 mixture of bivariate normals $\mathcal{N}((10, 10), \mathbf{\Sigma}_1)$ and $\mathcal{N}((-10, -10), \mathbf{\Sigma}_1)$ with $\mathbf{\Sigma}_1 = \begin{pmatrix} 10 & 3 \\ 3 & 2 \end{pmatrix}$.

Distributions S_2 and S_3 not only are centrally symmetric, but elliptically symmetric as well. This symmetry is more restrictive, which implies that it should be easier to detect the symmetry property in these cases. S_1 is the standard example of a centrally symmetric distribution which is not elliptically symmetric and S_4 is a somewhat artificial construction in which two identically distributed samples are centered in antipodal points. The symmetry of these two last distributions should be harder to identify.

Table 5.1 illustrates the confidence levels that were obtained from the simulations. It can be observed that the test succesfully identified all samples as having central symmetry. Even though not all levels correspond exactly to the nominal level of 5%, none of them are bellow this threshold, which means that the difference comes from a few samples in which the null hypothesis erroneously fails to be rejected. This type II error does not bring serious consequences in this case, since the distributions are indeed symmetric.

Table 5.1: Bootstrap levels for central symmetry testing with nominal level 5% for $d = 2$.

Distribution	$n = 100$	$n = 200$
S_1	3%	4%
S_2	4%	5%
S_3	5%	2%
S_4	0%	0%

To calculate the power of the test against alternatives we consider 6 different asymmetric distributions. Once again, when only a univariate distribution is given, it corresponds to the distribution of the marginals.

- A_1: A Weibull distribution with scale parameter $\lambda = 1$ and shape parameter $k = 1$.

- A_2: A Weibull distribution with scale parameter $\lambda = 1$ and shape parameter $k = 2$.

- A_3: A Weibull distribution with scale parameter $\lambda = 1$ and shape parameter $k = 2.5$.

- A_4: $exp(1)$.

- A_5: $\chi^2(10)$.

- A_6: A 40-60 mixture of normal distributions $\mathcal{N}((10, 10), \mathbf{\Sigma}_0)$ and $\mathcal{N}((-10, -10), \mathbf{\Sigma}_0)$.

- A_7: A 40-60 mixture of normal distributions $\mathcal{N}((1, 1), \mathbf{\Sigma}_0)$ and $\mathcal{N}((-1, -1), \mathbf{\Sigma}_0)$.

- A_8: A distribution designed to be angularly symmetric but not centrally symmetric. Vectors sampled from the uniform disrtibution on the unit circle are multiplied by its scalar projection onto $(1, 0)$. Thus the central symmetry is lost, but when projected onto the unit circle they are distributed uniformly and so have angular symmetry.

Distributions A_1 and A_4 are highly asymmetrical, they were used as a first test of the functionality of the method. A_2, A_3 and A_5 still look asymmetrical, but for some samples it is hard to tell from a plot, specially for A_3, since its shape parameter is close to the region in which the Weibull distribution is almost symmetrical. A_6 and A_7 where used to test the method for distributions whose symmetry properties are hard to distinguish, since they consist of subtle modifications of symmetric distributions such as S_4. In A_7 both normals get mixed and thus the difference in the number of points might get hidden, whereas in A_6 the centers are far and the probability of contact is minimal. Finally, A_8 was used to check that the test for angular symmetry worked for a distribution which is not centrally symmetric.

The results for $d = 2$ are shown in Table 5.2. The test shows good power against almost all alternatives and a general increase is observed when the sample size grows, most drastically for A_3. This indicates a desirable feature of the test: an improving performance with sample size. We can see low power against A_7, but the test still detects asymmetry, even when these alternatives were specially constructed to be almost symmetrical.

Table 5.2: Bootstrap power for central symmetry testing at 5% level for $d = 2$.

Alternative	$n = 100$	$n = 200$
A_1	0.92	0.96
A_2	0.87	0.91
A_3	0.45	0.81
A_4	1	1
A_5	0.98	0.99
A_6	0.99	1
A_7	0.13	0.13
A_8	0.87	1

For $d = 5$ we used the five dimensional distributions equivalent to S_1, S_2, S_3, A_1, A_2, A_3, A_4, A_5 and A_8, where the covariance matrix for the multivariate normal S_1 is

$$\mathbf{\Sigma} = \begin{pmatrix} 5 & 1 & 1 & 1 & 1 \\ 1 & 4 & 1 & 1 & 1 \\ 1 & 1 & 3 & 1 & 1 \\ 1 & 1 & 1 & 2 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix}.$$

S_4, A_6 and A_7 were not taken into account because in five dimesions they are not interesting examples to test the quality of the method.

The results for the simulation level for the symmetric distributions are given in Table 5.3 and the powers against alternatives are shown in Table 5.4. The results are similar to those with $d = 2$, with the test identifying symmetry and showing good power against all alternatives, except for A_3. However, it is important to notice that alternative A_3 showed the biggest increase of power with sample size for $d = 2$, hence the low power is probably a consequence of the reduced number of data. It is remarkable that despite considerably reducing the number of bootstraps and the sample size, the test still has the desired performance in dimension 5. In conclusion, our statistic seems to detect the lack of central symmetry of a distribution for different types of alternatives.

Table 5.3: Bootstrap levels for central symmetry testing with nominal level 5% for $d = 5$.

Distribution	$n = 50$
S_1	1%
S_2	0%
S_3	5%

Table 5.4: Bootstrap power for central symmetry testing at 5% level for $d = 5$.

Alternative	$n = 50$
A_1	1
A_2	0.65
A_3	0.14
A_4	1
A_5	0.9
A_8	0.55

5.2 Angular symmetry

The same conditions as for central symmetry were used. In this case the number of spherical harmonics for $d = 2$ is 4 and for $d = 5$ is 35. The same distributions were used to validate the test, with the difference that A_8 has angular symmetry so we rename it as S_5. Since angular symmetry can be viewed as central symmetry for the projections onto the unit sphere, it is to expect that the power against alternatives be lower. The results for $d = 2$ are given in Tables 5.5 and 5.6; for $d = 5$ we have Tables 5.7 and 5.8.

The results show lower powers against all alternatives, as was expected since angular symmetry is a relaxation of central symmetry. Nevertheless, the null hypothesis is rejected for all alternatives, and those of lowest power, A_2, A_3 and A_7, show a considerable increase in power with sample size, even though for $n = 200$ it is still very low. For S_5 the null hypothesis was not rejected, whereas it had been rejected by the test for central symmetry. This shows that the test seems to detect symmetry for distributions with no central symmetry. It is interesting to note that

Table 5.5: Bootstrap levels for angular symmetry testing with nominal level 5% for $d = 2$.

Distribution	$n = 100$	$n = 200$
S_1	5%	3%
S_2	2%	1%
S_3	5%	4%
S_4	3%	3%
S_5	1%	0%

Table 5.6: Bootstrap power for angular symmetry testing at 5% level for $d = 2$.

Alternative	$n = 100$	$n = 200$
A_1	0.74	0.97
A_2	0.13	0.19
A_3	0.05	0.09
A_4	0.78	0.99
A_5	0.46	0.44
A_6	0.98	1
A_7	0.06	0.1

Table 5.7: Bootstrap levels for angular symmetry testing with nominal level 5% for $d = 5$.

Distribution	$n = 50$
S_1	3%
S_2	1%
S_3	4%
S_5	3%

alternatives A_2 and A_4 have higher power in dimension 5 even when the sample size is half of that in $d = 2$; this fact could indicate that the small number of spherical harmonics for $d = 2$ affects the efficiency of the test.

To conclude, both tests seems to detect departure from symmetry for different alternatives, with the test for central symmetry being considerably more efficient. To improve the results of the test for angular symmetry one may consider spherical

Table 5.8: Bootstrap power for angular symmetry testing at 5% level for $d = 5$.

Alternative	$n = 50$
A_1	0.77
A_2	0.25
A_3	0.13
A_4	0.85
A_5	0.38

harmonics of higher order or another appropiate class of functions. However, the lower efficiency with respect to the test of central symmetry is intrinsic to the relationship between these definitions and thus one cannot expect to equalize or surpass the results of the latter test. Consequently, if a distribution is suspected to have both kinds of symmetry it is better to test for central symmetry and only in case of failure apply the angular symmetry test.

Chapter 6

Conclusions and prospects

6.1 Conclusions

- The bootstrap method based on the empirical distribution allowed the development of tests for central and angular symmetry of simple implementation.

- The proposed tests seems to detect departure from the respective symmetries for different types of alternatives. In particular, a distribution which is angularly symmetric but not centrally symmetric was correctly identified by the tests.

- Multivariate Legendre polynomials and spherical harmonics work as test functions even for small orders.

- Tests for central and angular symmetry which are based on the same method will have more efficiency detecting departure from central symmetry due to the inclusion relationship between these definitions. In consequence, if a distribution is suspected to have both kinds of symmetry it is recommended to test for central symmetry and only in case of failure, to apply the angular symmetry test.

6.2 Prospects

- The extension of the method to other types of symmetry can be studied. In particular, it would be challenging to adapt the ideas of the method to a symmetry given by an infinite group, such as spherical or elliptical symmetry.

- The tests for central and angular symmetry might be improved by considering other classes of functions, not necessarily polynomials.

- It would be valuable to make a comparison study between existing tests for central and angular symmetry.

Appendix A

Some probability theory definitions.

Let $(\Omega, \mathcal{A}, P)$ be a probability space.

Definition A.1. The *outer probability* of a subset $B \subseteq \Omega$ is

$$P^*(B) = \inf\{P(A) : A \supset B, A \in \mathcal{A}\}.$$

Definition A.2. Let $T : \Omega \mapsto \bar{\mathbb{R}}$ be an arbitrary map ($\bar{\mathbb{R}}$ is the extended real line $[-\infty, \infty]$). The *outer integral* of T with respect to P is given by

$$\mathrm{E}^*T = \inf\{\mathrm{E}U : \ U \geqslant T, U : \Omega \mapsto \mathbb{R} \text{ measurable and } \mathrm{E}U \text{ exists}\}.$$

Definition A.3. Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of arbitrary, possibly nonmeasurable, maps from underlying probability spaces $(\Omega_n, \mathcal{A}_n, P_n)$ to a metric space $\mathbb{D}$. The sequence $(X_n)_{n\in\mathbb{N}}$ converges weakly to a Borel measurable map X, written $X_n \rightsquigarrow X$, if

$$\mathrm{E}^* f(X_n) \to \mathrm{E} f(X), \text{ for every} f \in C_b(\mathbb{D}),$$

where $C_b(\mathbb{D})$ is the space of all bounded, continuous, real functions on $\mathbb{D}$.

Definition A.4. An estimator θ is called $n^{1/2}$-consistent if it is asymptotically normal.

Definition A.5. The following are the smoothness conditions (S) required for P and $\mathcal{F}$ in Chapter 4.

- P is absolutely continuous with a uniformly bounded and continuously differentiable density p such that for some $C_A > d+1$

$$\sup_{x \in \mathbb{R}^d} (1+|x|)^{C_A} |p'(x)| < \infty.$$

- The class $\mathcal{F}$ is uniformly bounded and for all $\epsilon > 0$ and $R > 0$

$$\sup_{f \in \mathcal{F}} \lambda^d \{x \in \mathbb{R}^d : |x| \leq R \text{ and } \omega_f(x;\delta) \geq \epsilon\} \to 0 \text{ as } \delta \to 0,$$

  where λ^d represents the Lebesgue measure in $\mathbb{R}^d$ and

$$\omega_f(x;\delta) := \sup\{|f(x_1) - f(x_2)| : |x_1 - x| \leq \delta, |x_1 - x| \leq \delta\}$$

Bibliography

T. W. Anderson and K.-T. Fang. *Statistical Inference in Elliptically Contoured and Related Distributions.* Allerton Press, New York, 1990.

M. Arcones and E. Giné. Some bootstrap tests for symmetry for univariate distributions. *Ann. Statist.*, 19:1496–1511, 1991.

L. Baringhaus. Testing for spherical symmetry of a multivariate distribution. *Ann. Stat.*, 19:899–917, 1991.

R. J. Beran. Testing for ellipsoidal symmetry of a multivariate distribution. *Ann. Stat.*, 7:150–162, 1979.

D. K. Blough. Multivariate symmetry via projection pursuit. *Ann. Inst. Stat. Math.*, 41:461–475, 1989.

A. P. Dempster. *Elements of Continuous Multivariate Analysis.* Addison-Wesley, New York, 1969.

C. Dicks and H. Tong. A test for symmetries of multivariate probability distributions. *Biometrika*, 86:605–614, 1999.

K.-T. Fang, R.-Z. Li, and L.-X. Zhu. A projection nt-type test of ellyptical symmetry based on the skewness and kurtosis measures. *Acta Math. Appl. Sin. (English series)*, 14:314–323, 1995.

S. Ghosh and F. H. Ruymgaart. Projection pursuit type applications of empirical characteristic functions in some multivariate problems. *Can. J. Stat.*, 20:429–440, 1992.

E. Giné and J. Zinn. Gaussian characterization of uniform donsker classes. *Ann. Probab.*, 19:758–782, 1991.

J. M. González-Barrios and R. Rueda. On convergence theorems for quantiles. *Commun. Stat. Theory Meth.*, 30:943–944, 2001.

A. K. Gupta and T. Varga. Multivariate robust tests for spherical symmetry with applications to multivariate least squares regression. *J. Appl. Stat. Sci.*, 1:159–168, 1993.

T. Kariya and M. L. Eaton. Robust tests for spherical symmetry. *Ann. Stat.*, 54: 206–215, 1977.

V. Koltchinskii and L. Li. Testing for spherical symmetry of a multivariate distribution. *J. Multivariate Anal.*, 65:228–244, 2000.

V. Koltchinskii and L. Sakhanenko. Testing for ellipsoidal symmetry of a multivariate distribution. In E. Giné, D. Mason, and J. Wellner, editors, *High Dimensional Probability II*, pages 493–510. Birkhäuser, Boston, 2000.

S. Kotz, N. Balakrishnan, and N. L. Johnson. *Continuous Multivariate Distributions, Volume 1: Models and Applications.* John Wiley & Sons, New York, 2000.

R.-Z. Li, K.-T. Fang, and L.-X. Zhu. Some q-q probability plots to test spherical and ellyptical symmetry. *J. Comput. Graph. Stat.*, 6:435–450, 2001.

R. Y. Liu, J. M. Parelius, and K. Singh. Multivariate analysis by data depth: Descriptive statistics, graphics and inference (with discussion). *Ann. Stat.*, 27: 783–858, 1999.

A. Manzotti and A. J. Quiroz. Spherical harmonics in quadratic forms for testing multivariate normality. *Sociedad de Estadística e Investigación Operativa Test*, 2001.

A. Manzotti, F. J. Pérez, and A. J. Quiroz. A statistic for testing the null hypothesis of elliptical symmetry. *J. Multivariate Anal.*, 81:274–285, 2002.

M. Petitjean. Chirality and symmetry measures: a transdisciplinary review. *Entropy*, 5:271–312, 2003.

David Pollard. *Convergence of Stochastic Processes.* Springer-Verlag, New York, 1984.

A. J. Quiroz and R. M. Dudley. *Probability Theory and Related Fields*, 87.

J. Romano. Bootstrap and randomization tests of some nonparametric hypothesis. *Ann. Statist.*, 17:141–159, 1989.

L. Sakhanenko. Testing group symmetry of a multivariate distribution. *Symmetry*, 1(2):180–200, 2009.

R. J. Serfling. *Multivariate symmetry and assymetry.* Technical Report, University of Texas, 2006.

G. Shorack and J. Wellner. *Empirical Processes with Applications to Statistics.* Wiley, New York, 1986.

A. W. van der Vaart. *Asymptotic Statistics.* Cambridge University Press, 2000.

A. W. van der Vaart and J. A. Wellner. *Weak Convergence and Empirical Processes with Applications to Statistics.* Springer-Verlag, New York, 1996.

T. Zheng and J. L. Gastwirth. On bootstrap tests of symmetry about an unknown median. *Journal of Data Science*, 8:397–412, 2010.

L.-X. Zhu and G. Neuhaus. Nonparametric monte carlo tests for multivariate distributions. *Biometrika*, 87:919–928, 2000.

L.-X. Zhu, K.-T. Fang, and J.-T. Zhang. A projection nt-type test for spherical symmetry of a multivariate distribution. In E.-M. Tiit, T. Kollo, and H. Niemi., editors, *New Trends in Probability and Statistics, Volume 3: Multivariate Statistics and Matrices in Statistics.*, pages 109–122. eds. VSP, Utrecht, 1995.

Y. Zuo and R. Serfling. On the performance of some robust nonparametric location measures relative to a general notion of multivariate symmetry. *J. Stat. Plann. Inference*, 84:55–79, 2000.