UNIVERSITY OF BELGRADE
FACULTY OF MATHEMATICS

Katarina Halaj Mileusnić

Novel goodness-of-fit tests based on independence-type characterizations

Doctoral Dissertation

Belgrade, 2026

УНИВЕРЗИТЕТ У БЕОГРАДУ
МАТЕМАТИЧКИ ФАКУЛТЕТ

Катарина Халај Милеуснић

Нови тестови сагласности засновани на карактеризацијама независности статистика

Докторска дисертација

Београд, 2026.

Advisor:

Professor Bojana Milošević, PhD, Associate Professor
Faculty of Mathematics, University of Belgrade
Chair of the Department of Probability and Statistics

Dissertation defense committee members:

Professor Marko Obradović, PhD, Associate Professor
Faculty of Mathematics, University of Belgrade

Marija Cuparić, PhD, Assistant Professor
Faculty of Mathematics, University of Belgrade

Professor María Dolores Jiménez-Gamero, PhD, Full Professor
Dpto. Estadística e Investigación Operativa, University of Seville

Acknowledgements

Reaching the end of this journey has made me realize how many people have contributed to it through their support, guidance, encouragement, and understanding. Although words can hardly convey the full extent of my gratitude, I would nevertheless like to try to express my heartfelt thanks to everyone who stood by me along the way.

First, I would like to express my sincere gratitude to my supervisor, Professor Bojana Milošević, for her unwavering support, patience, and invaluable guidance, especially when the challenges seemed overwhelming - which, admittedly, happened more often than I would like to confess.

I am also deeply grateful to the members of my dissertation committee, Prof. Dr. María Dolores Jiménez-Gamero, Dr. Marija Ćuparić and Prof. Dr. Marko Obradović, for their careful reading of the manuscript and their valuable suggestions and comments, which greatly contributed to its improvement.

I owe special thanks to my colleagues from Office 114 and to my colleagues from the Department of General and Applied Mathematics, who were always there to listen and offer their support, and whose humour made the Faculty a place I was always happy to come to.

Beyond the academic environment, I am deeply grateful to my family and friends, whose love, patience, understanding, and encouragement accompanied me throughout this journey. Their presence gave me strength during difficult moments and reminded me that there was always something to look forward to beyond the pages of this dissertation.

My deepest gratitude goes to those who were my greatest support throughout this journey.

Mum, thank you for showing me, through your strength and determination, that giving up was never even the slightest possibility. Dad, thank you for always understanding me without the need for words, and for making me smile when I needed it most. Deki, thank you for helping me grow stronger and become a true fighter - though not a runner. Tara, thank you for teaching me what it means to love unconditionally. Finally, Petar, I am profoundly grateful to you for never ceasing to believe in me, and for making every day better simply by being there.

Without all of you, this work would not have been possible.

To my family, with all my love and gratitude.

Изјава захвалности

Тек на крају овог пута постала сам свесна колико је људи својом подршком, саветима, охрабрењем и разумевањем оставило траг на њему. Иако речи не могу да обухвате сву захвалност коју осећам, желела бих да бар покушам да је искажем и захвалим се онима који су били уз мене на сваком кораку.

Најпре дугујем искрену захвалност својој менторки, проф. др Бојани Милошевић, на несебичној подршци, стрпљењу и драгоценим саветима током израде ове дисертације, нарочито у тренуцима када су изазови деловали непремостиво - што се, додуше, дешавало чешће него што бих желела да признам.

Велику захвалност дугујем и члановима Комисије, проф. др Марији Долорес Хименес-Гамеро, др Марији Цупарић и проф. др Марку Обрадовићу, на пажљивом читању рукописа, корисним сугестијама и конструктивним коментарима који су значајно допринели унапређењу овог рада.

Посебно се захваљујем колегама из канцеларије 114, као и колегама са Катедре за општу и примењену математику, који су увек били ту да саслушају и пруже подршку и који су својим хумором чинили Факултет местом на које сам увек радо долазила.

Изван академског окружења, неизмерно сам захвална својој породици и пријатељима чија су ме љубав, стрпљење, разумевање и охрабривање пратили током целог овог пута. Њихово присуство давало ми је снагу у тренуцима када је било најтеже и подсећало ме да је увек постојало нешто чему сам могла да се радујем изван страница ове дисертације.

Моја највећа захвалност припада онима који су ми током свих ових година били највећи ослонац.

Мама, хвала ти што си ми својом снагом и одлучношћу показала да одустајање никада није било ни најмања опција. Тата, хвала ти што си ме увек разумео и без много речи и што си умео да ми измамиш осмех онда када ми је био најпотребнији. Деки, хвала ти што си ми помогао да ојачам и постанем прави борац - мада не и тркач. Тара, хвала ти што си ме научила шта значи волети безусловно. И на крају, Петре, бескрајно ти хвала што ни у једном тренутку ниси престао да верујеш у мене и што сваки мој дан чиниш лепшим.

Без свих вас овај рад не би био могућ.

Мојој породици, с љубављу и неизмерном захвалношћу.

Dissertation title: Novel goodness-of-fit tests based on independence-type characterizations

Abstract: The main objective of this dissertation is to develop novel goodness-of-fit tests based on independence characterizations and to investigate their properties. The starting point is the fact that certain distributions can be uniquely characterized by the independence of suitable functions of random variables. This observation makes it possible to reduce goodness-of-fit testing for a hypothesized model to the problem of detecting departures from the corresponding independence condition. Accordingly, the proposed test statistics are constructed by comparing empirical versions of joint functionals with the product of the corresponding marginal functionals, where the resulting discrepancy serves as the basis for new goodness-of-fit tests for different classes of distributions.

One of the contributions of the dissertation is the development of goodness-of-fit tests for the geometric distribution based on Ferguson's independence characterization. In this setting, the general framework of comparing joint and marginal empirical functionals is implemented through V-empirical probability generating functions. Furthermore, a novel approach to goodness-of-fit testing for absolutely continuous distributions is proposed, relying on the discrepancy between the joint U-empirical distribution function of an appropriately defined random vector and the product of its marginal U-empirical distribution functions. Another part of the research focuses on circular data. A goodness-of-fit test for the wrapped normal distribution is proposed, where the general methodology is adapted to the periodic structure of observations on the circle.

The theoretical part of the dissertation is devoted to the derivation of the asymptotic properties of the proposed test statistics, drawing on the theory of U- and V- statistics as well as on the theory of U-empirical processes. An additional contribution is the establishment of new asymptotic results for finite linear combinations of weakly degenerate V-statistics, both in settings with known parameters and in settings involving parameter estimation. These findings provide the theoretical foundation for a class of combined tests in which appropriately chosen weights regulate the contribution of individual components and can be used to enhance the overall efficiency of the testing procedure.

The practical performance of the proposed methods will be assessed through extensive simulation studies and the analysis of several real data sets. The simulation study will investigate the finite-sample properties of the tests under the null hypothesis and a range of relevant alternatives, while the real-data applications will demonstrate their usefulness in practical statistical problems. In this way, the empirical part of the dissertation will complement the theoretical developments and provide evidence of the practical applicability of the proposed methodologies.

Keywords: goodness-of-fit testing, independence characterizations, linear combinations of degenerate V-statistics, circular distributions, bootstrap

Scientific Area: Mathematics

Scientific Sub-Area: Probability and Statistics

Наслов дисертације: Нови тестови сагласности засновани на карактеризацијама независности статистика

Резиме: Главни циљ ове дисертације јесте развој нових тестова сагласности заснованих на карактеризацијама независности, као и испитивање њихових својстава. Полазна идеја је да се поједине расподеле могу једнозначно одредити условом независности одговарајућих функција случајних променљивих. То омогућава да се провера сагласности са претпостављеним моделом сведе на испитивање одступања од тог услова независности. У складу са тим, предложене тест статистике заснивају се на поређењу емпиријских верзија заједничких функционала са производом одговарајућих маргиналних функционала, при чему се њихова разлика користи као основа за конструкцију нових тестова сагласности за различите класе расподела.

Један од доприноса дисертације јесте развој тестова сагласности за геометријску расподелу, заснованих на Фергусоновој карактеризацији независности. У овом случају општи приступ заснован на поређењу заједничких и маргиналних емпиријских функционала реализује се преко V-емпиријских вероватносних генераторних функција. Поред тога, предложен је нови приступ тестирању сагласности за апсолутно непрекидне расподеле, заснован на разлици заједничке U-емпиријске функције расподеле одговарајућег случајног вектора и производа њених маргиналних U-емпиријских функција расподеле. Још један од праваца истраживања посвећен је циркуларним подацима. Предложен је одговарајући тест сагласности за обавијену нормалну расподелу, уз прилагођавање опште методологије периодичној структури опсервација на кругу.

Теоријски део дисертације усмерен је на извођење асимптотских својстава предложених тест-статистика, уз ослањање на теорију U- и V- статистика, као и теорију U-емпиријских процеса. Додатни допринос представља извођење нових резултата о асимптотским расподелама коначних линеарних комбинација слабо дегенерисаних V- статистика, како у случају познатих параметара, тако и у ситуацијама у којима се параметри оцењују на основу података. Добијени резултати служе као теоријска основа за креирање комбинованих тестова, код којих се избором тежина контролише утицај појединачних компоненти и настоји побољшати њихова укупна ефикасност.

Практично понашање предложених метода испитује се кроз симулационе студије и анализу неколико реалних скупова података. Приликом симулација разматрају се коначноузорачке особине тестова под нултом хипотезом и изабраним алтернативама, док примене на реалним подацима илуструју њихову употребу у конкретним статистичким проблемима. На тај начин емпиријски део дисертације допуњује теоријске резултате и илуструје практичну применљивост предложених метода.

Кључне речи: тестови сагласности, карактеризације независности, линеарне комбинације дегенерисаних V-статистика, циркуларне расподеле, бутстреп

Научна област: Математика

Ужа научна област: Вероватноћа и статистика

Contents

List of Abbreviations

Abbreviation	Meaning
gof	goodness-of-fit
pmf	probability mass function
i.i.d.	independent and identically distributed
iff	if and only if
pgf	probability generating function
edf	empirical distribution function
df	distribution function
pdf	probability density function
cdf	cumulative distribution function
mle	maximum likehood estimator

List of Symbols

Symbol	Meaning
$\mathbb{N} = \{1, 2, \dots\}$	The set of natural numbers
$\mathbb{Z}$	The set of integer numbers
$\mathbb{R}$	The set of real numbers
$\overline{\mathbb{R}}^2$	The extended real plane $[-\infty, \infty]^2$
$\xrightarrow{\mathcal{D}}$	Convergence in distribution
$\xrightarrow{\mathbb{P}}$	Convergence in probability
$\xrightarrow{a.s.}$	Almost sure convergence
$\mathbb{E}(X)$	Expected value of the random variable X
$\mathrm{Var}(X)$	Variance of the random variable X
$\mathrm{Cov}(X, Y)$	Covariance of the random variables X and Y
$\mathcal{N}(\mu, \sigma^2)$	Univariate normal dist. with mean μ and variance σ^2
$\mathcal{N}_d(\boldsymbol{\mu}, \Sigma)$	d-variate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix Σ
χ^2_a	Chi-squared distribution with a degrees of freedom
$\langle \cdot, \cdot \rangle$	Inner product
$\mathbf{1}\{A\}$	Indicator of the event A, 1 if A occurs and 0 otherwise
δ_x	Dirac measure concentrated at x
$\binom{a}{b}$	Binomial coefficient
$f_n = O(g_n)$	$f_n = h_n g_n$ for some bounded sequence (h_n)
$f_n = o(g_n)$	$f_n = h_n g_n$, where $h_n \to 0$ as $n \to \infty$
$f_n = O_{\mathbb{P}}(g_n)$	$f_n = h_n g_n$ for some seq. (h_n) bounded in probability
$f_n = o_{\mathbb{P}}(g_n)$	$f_n = h_n g_n$, where $h_n \xrightarrow{\mathbb{P}} 0$ as $n \to \infty$
$\mathrm{tr}(A)$	Trace of the operator A
$\lfloor x \rfloor$	Greatest integer less than or equal to x
$\mathrm{sgn}(x)$	Sign function: -1 if $x < 0$, 0 if $x = 0$, and 1 if $x > 0$
$a \equiv b \pmod{2\pi}$	a and b differ by an integer multiple of 2π
$\bar{z}$	Complex conjugate of the complex number z
$\cosh(x)$	Hyperbolic cosine of x
$\sinh(x)$	Hyperbolic sine of x
$\Gamma(x)$	Gamma function evaluated at x

Chapter 1
Introduction

A common approach to constructing goodness-of-fit tests is to begin with a property that uniquely identifies the distribution under the null hypothesis. In characterization-based testing, such a property is valid only under the null hypothesis. The test statistic is then designed to measure the departure from that property. In this dissertation, attention is restricted to characterizations in which the defining property is expressed through the independence of suitably constructed random quantities. The passage from the theoretical characterization to its empirical version leads to test statistics whose analysis relies on the theory of U- and V-statistics and their process analogues. Therefore, a careful study of the corresponding asymptotic behavior is required.

Accordingly, this introductory chapter presents the theoretical tools needed for the analysis carried out in the rest of the dissertation. It begins with a brief overview of selected convergence results and limit theorems for random elements taking values in metric and Hilbert spaces. Afterwards, the basic theory of U- and V-statistics is presented, including their definitions, illustrative examples, and fundamental results concerning their limiting distributions. This theory is essential for the dissertation, since the test statistics studied in the subsequent chapters are formulated as U- and V-statistics or their functionals. The well-established results presented in this chapter can be found in Korolyuk and Borovskikh (2013), Serfling (1980) and Billingsley (1999) as well as in the references cited throughout the chapter. This chapter also contains novel results concerning finite linear combinations of degenerate V-statistics, both in the known-parameter and unknown-parameter settings (see Theorem 1.12, Corollary 1.1 and 1.2). The chapter then proceeds to U-processes, which arise as a natural extension of U-statistics and play an important role in the subsequent construction and analysis of test statistics. The last part of the chapter introduces distributional characterizations in general and then focuses on independence characterizations, which are the basis for the gof tests proposed in the thesis.

1.1 Convergence results for random elements in metric and Hilbert spaces

The asymptotic analysis of the test statistics considered in this dissertation involves empirical processes and related statistic-based quantities as random elements in suitable function spaces. Since the convergence results used later are naturally formulated in the general setting of metric spaces, this section begins with the corresponding basic notions and results. These results will then be applied, in particular, to random elements taking values in Hilbert spaces.

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and let $(\mathcal{M}, d)$ be a metric space equipped with its Borel σ-field $\mathcal{B}(\mathcal{M})$. A mapping

$$X : \Omega \to \mathcal{M}$$

is called a random element in $\mathcal{M}$ if $X^{-1}(B) \in \mathcal{A}$, $\forall B \in \mathcal{B}(\mathcal{M})$.

Special cases of this definition include the usual notions of random variables and random vectors. Namely, if $\mathcal{M} = \mathbb{R}$, then X is a random variable, while for $\mathcal{M} = \mathbb{R}^k$, X is a random vector. If $\mathcal{M}$ is a function space, then X is referred to as a random function.

The distribution of X is the probability measure P_X on $(\mathcal{M}, \mathcal{B}(\mathcal{M}))$ defined by

$$P_X(A) = \mathbb{P}(X \in A) = \mathbb{P}\{\omega \in \Omega : X(\omega) \in A\}, \quad A \in \mathcal{B}(\mathcal{M}).$$

Let P and P_n, $n \geq 1$, be probability measures on $(\mathcal{M}, \mathcal{B}(\mathcal{M}))$. Weak convergence of P_n to P, denoted by

$$P_n \Rightarrow P, \quad n \to \infty$$

means that

$$\int_{\mathcal{M}} f \, dP_n \longrightarrow \int_{\mathcal{M}} f \, dP, \quad n \to \infty$$

for every bounded continuous function $f : \mathcal{M} \to \mathbb{R}$.

Let $X, X_1, X_2, \ldots$ be random elements in $\mathcal{M}$, with corresponding distributions $P_X, P_{X_1}, P_{X_2}, \ldots$. Convergence in distribution of X_n to X is defined as weak convergence of the corresponding distributions, that is,

$$P_{X_n} \Rightarrow P_X, \quad n \to \infty.$$

This convergence is also denoted by

$$X_n \xrightarrow[n\to\infty]{\mathcal{D}} X.$$

In addition to convergence in distribution, two stronger modes of convergence will also be used. The sequence $\{X_n\}_{n\geq 1}$ is said to converge in probability to X if, for every $\varepsilon > 0$,

$$\mathbb{P}\left(d(X_n, X) > \varepsilon\right) \longrightarrow 0, \quad n \to \infty.$$

This convergence is denoted by

$$X_n \xrightarrow[n\to\infty]{\mathbb{P}} X.$$

The sequence $\{X_n\}_{n\geq 1}$ is said to converge almost surely to X if

$$\mathbb{P}\left(\left\{\omega \in \Omega : \lim_{n\to\infty} d(X_n(\omega), X(\omega)) = 0\right\}\right) = 1.$$

This convergence is denoted by

$$X_n \xrightarrow[n\to\infty]{a.s.} X.$$

A set $A \in \mathcal{B}(\mathcal{M})$ is called a P-continuity set if its boundary ∂A satisfies

$$P(\partial A) = 0.$$

Since ∂A is closed, it belongs to $\mathcal{B}(\mathcal{M})$.

The following result gives several equivalent formulations of weak convergence of probability measures.

Lemma 1.1 (Portmanteau lemma, Billingsley, 1999)**.** *Let $(\mathcal{M}, d)$ be a metric space, and let P_n, $n \geq 1$, and P be probability measures on $(\mathcal{M}, \mathcal{B}(\mathcal{M}))$. Then the following conditions are equivalent:*

(i) $P_n \Rightarrow P, \quad n \to \infty.$

(ii) For every bounded uniformly continuous function $f : \mathcal{M} \to \mathbb{R}$,

$$\int_{\mathcal{M}} f \, dP_n \longrightarrow \int_{\mathcal{M}} f \, dP, \quad n \to \infty.$$

(iii) For every closed set $F \subset \mathcal{M}$,

$$\limsup_{n\to\infty} P_n(F) \leq P(F).$$

(iv) For every open set $G \subset \mathcal{M}$,

$$\liminf_{n\to\infty} P_n(G) \geq P(G).$$

(v) For every P-continuity set $A \in \mathcal{B}(\mathcal{M})$,

$$\lim_{n\to\infty} P_n(A) = P(A).$$

Since the next result is formulated for random elements taking values in separable metric spaces, separability of metric spaces is briefly recalled before its statement.

Definition 1.1. *A metric space $\mathcal{M}$ is called separable if there exists a countable subset $D = \{x_1, x_2, \dots\} \subset \mathcal{M}$ such that $\overline{D} = \mathcal{M}$. Equivalently, for every $h \in \mathcal{M}$ and every $\varepsilon > 0$, there exists $x_k \in D$ satisfying*

$$d(h, x_k) < \varepsilon.$$

For random elements $X_1, \dots, X_n$ taking values in $\mathcal{M}$, the empirical measure is defined by

$$\mu_n(B) = \frac{1}{n}\sum_{j=1}^{n} \delta_{X_j}(B) = \frac{1}{n}\sum_{j=1}^{n} \mathbf{1}\{X_j \in B\}, \quad \text{for } B \in \mathcal{B}(\mathcal{M}).$$

The next theorem provides a general consistency result for empirical measures. It states that for independent random elements taking values in a separable metric space, the empirical measure converges weakly almost surely to the underlying probability measure.

Theorem 1.1 (Varadarajan theorem, Dudley, 2002)**.** *Suppose that $(\mathcal{M}, d)$ is a separable metric space and that μ is a Borel probability measure on $\mathcal{M}$. Let $X_1, X_2, \dots$ be independent random elements in $\mathcal{M}$ with common distribution μ, and let*

$$\mu_n = \frac{1}{n}\sum_{j=1}^{n} \delta_{X_j}$$

denote the empirical measure. Then the empirical measures μ_n converge a.s. to μ, i.e.

$$\mathbb{P}\left(\{\omega : \mu_n(\cdot)(\omega) \to \mu\}\right) = 1.$$

The convergence of random elements is frequently used as a starting point for deriving the limiting distributions of statistics obtained as their functionals. The passage from the convergence of the underlying random elements to the convergence of the corresponding statistics is justified by the continuous mapping theorem stated below, provided that the functional is measurable and continuous at the limiting element with probability one; see, for example, van der Vaart, 1998; Billingsley, 1999.

Theorem 1.2 (Continuous mapping theorem)**.** *Let $(\mathcal{M}, d)$ and $(\mathcal{M}', d')$ be metric spaces, and let X_n and X be random elements taking values in $\mathcal{M}$. Let $g : \mathcal{M} \to \mathcal{M}'$ be a Borel measurable function, and denote by D_g the set of discontinuity points of g, such that $\mathbb{P}(X \in D_g) = 0$. Then the following implications hold:*

(i) If $X_n \xrightarrow[n\to\infty]{a.s.} X$, then $g(X_n) \xrightarrow[n\to\infty]{a.s.} g(X)$.

(ii) If $X_n \xrightarrow[n\to\infty]{\mathbb{P}} X$, then $g(X_n) \xrightarrow[n\to\infty]{\mathbb{P}} g(X)$.

(iii) If $X_n \xrightarrow[n\to\infty]{\mathcal{D}} X$, then $g(X_n) \xrightarrow[n\to\infty]{\mathcal{D}} g(X)$.

The continuous mapping theorem is often used together with Slutsky's theorem in the derivation of limiting distributions. Slutsky's theorem is particularly useful when a weakly convergent sequence has to be combined with another sequence converging in probability to a non-random limit. For this reason, the version of Slutsky's theorem for random elements in metric spaces is stated below, in the form that will be used throughout the dissertation (see, e.g., van der Vaart and Wellner, 1996).

Theorem 1.3 (Slutsky's theorem)**.** *Let $(\mathcal{X}, d_x)$ and $(\mathcal{Y}, d_y)$ be metric spaces. Let X_n and X be $\mathcal{X}$-valued Borel measurable random elements, and let Y_n be $\mathcal{Y}$-valued Borel measurable random elements. Assume that X is separable, that is, there exists a separable subset $\mathcal{X}_0 \subset \mathcal{X}$ such that*

$$\mathbb{P}(X \in \mathcal{X}_0) = 1.$$

If

$$X_n \xrightarrow[n\to\infty]{\mathcal{D}} X \quad \text{in } \mathcal{X}$$

and

$$Y_n \xrightarrow[n\to\infty]{\mathcal{D}} y \quad \text{in } \mathcal{Y},$$

where $y \in \mathcal{Y}$ is non-random, then

$$(X_n, Y_n) \xrightarrow[n\to\infty]{\mathcal{D}} (X, y) \quad \text{in } \mathcal{X} \times \mathcal{Y},$$

where $\mathcal{X} \times \mathcal{Y}$ is equipped with a product metric.

Since every Hilbert space is a metric space with respect to the metric induced by its norm, the above theorems apply directly to random elements taking values in Hilbert spaces.

Throughout the remainder of this section, let $\mathcal{H}$ be a Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and induced norm

$$\|x\| = \sqrt{\langle x, x \rangle}, \quad x \in \mathcal{H}.$$

A random element X taking values in $\mathcal{H}$ is understood as a Borel measurable mapping into $\mathcal{H}$. Within this framework, the central limit theorem for Hilbert-valued random elements can be stated in the following form (see, e.g., van der Vaart and Wellner, 1996).

Theorem 1.4 (Central limit theorem)**.** *If $X_1, X_2, \ldots$ are i.i.d. Borel measurable random elements in a separable Hilbert space $\mathcal{H}$ such that*

$$\mathbb{E}\langle X_1, h \rangle = 0 \quad \text{for every } h \in \mathcal{H},$$

and

$$\mathbb{E}\|X_1\|^2 < \infty.$$

Then the sequence

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i$$

converges in distribution to a Gaussian variable G. The distribution of G is determined by the distribution of its marginals $\langle G, h \rangle$, which are $\mathcal{N}(0, \mathbb{E}\langle X_1, h \rangle^2)$-distributed for every $h \in \mathcal{H}$.

For the analysis of bootstrap approximations, limit theorems for triangular arrays of Hilbert-space-valued random elements are required. In particular, results of this type allow one to establish the conditional weak convergence of bootstrap versions of the statistics under consideration.

Theorem 1.5 (Kundu et al., 2000)**.** *Let $\mathcal{H}$ be a separable Hilbert space and let $\{e_k : k \geq 1\}$ be an orthonormal basis of $\mathcal{H}$. For each $n \in \mathbb{N}$, let $Y_{n1}, Y_{n2}, \ldots, Y_{nn}$ be independent*

$\mathcal{H}$-valued random elements such that

$$\mathbb{E}\big(\langle Y_{ni}, e_k\rangle\big) = 0 \quad \textit{and} \qquad \mathbb{E}\big(\|Y_{ni}\|^2\big) < \infty$$

for all $i = 1, \ldots, n$ and $k \geq 1$. Denote by

$$S_n := \sum_{i=1}^{n} Y_{ni}$$

the corresponding partial sum, and let Γ_n denote the covariance operator of S_n. Assume that the following conditions are satisfied:

(i) For every $k, \ell \geq 1$,

$$\lim_{n\to\infty} \langle \Gamma_n e_k, e_\ell\rangle = \alpha_{k\ell}.$$

(ii)

$$\lim_{n\to\infty} \sum_{k=1}^{\infty} \langle \Gamma_n e_k, e_k\rangle = \sum_{k=1}^{\infty} \alpha_{kk} < \infty.$$

(iii) For every $\varepsilon > 0$ and every $k \geq 1$,

$$\lim_{n\to\infty} L_n(\varepsilon, e_k) = 0,$$

where for every $h \in \mathcal{H}$

$$L_n(\varepsilon, h) := \sum_{i=1}^{n} \mathbb{E}\left(\langle Y_{ni}, h\rangle^2 \mathbf{1}\{|\langle Y_{ni}, h\rangle| > \varepsilon\}\right).$$

Then, S_n converge in distribution to the centered Gaussian measure on $\mathcal{H}$ with covariance operator Γ, determined by

$$\langle \Gamma h, e_k\rangle = \sum_{\ell=1}^{\infty} \langle h, e_\ell\rangle \alpha_{\ell k}, \quad k \geq 1, \quad h \in \mathcal{H}.$$

1.2 U- and V- statistics

Consider a nonparametric family $\mathcal{P}$ of probability measures defined on an arbitrary measurable space. For $F \in \mathcal{P}$, let $\theta(F)$ be a real-valued functional.

Definition 1.2. *A parameter $\theta(F)$ is said to be an estimable parameter within $\mathcal{P}$ if there exists an integer m and a real-valued measurable function $\Phi(x_1, \ldots, x_m)$ such that, for*

i.i.d. random variables $X_1, \ldots, X_m$ with distribution F and support $S \subset \mathbb{R}^d$,

$$\mathbb{E}_F\big[\Phi(X_1, \ldots, X_m)\big] = \theta(F) \quad \textit{for all } F \in \mathcal{P}.$$

The smallest such integer m is called the degree of $\theta(F)$.

It should be noted that the function Φ can always be taken to be symmetric in its arguments. This is because, given any unbiased estimator Ψ of $\theta(F)$, one can form a new function by averaging Ψ over all possible permutations of its input variables. The resulting function is symmetric by construction and retains the same expectation under any distribution $F \in \mathcal{P}$ as the original estimator Ψ. In other words, symmetrization does not affect unbiasedness, but ensures that the estimator treats all its arguments equivalently.

Definition 1.3. *Let $\Phi(x_1, \ldots, x_m)$ be a real-valued measurable function symmetric in its arguments, and let $X_1, \ldots, X_n$ be a sample of size $n \geq m$ with a distribution F and support $S \subset \mathbb{R}^d$. A U-statistic with kernel Φ of degree m is defined as*

$$U_n = U_n(\Phi) = \frac{1}{\binom{n}{m}} \sum_{(i_1,\ldots,i_m)\in C_{m,n}} \Phi(X_{i_1}, \ldots, X_{i_m}),$$

where the summation is over the set $C_{m,n}$ of increasing m-tuples of indices from $\{1, \ldots, n\}$ with distinct entries.

If $\theta(F) = \mathbb{E}_F[\Phi(X_1, \ldots, X_m)]$ exists for all $F \in \mathcal{P}$, then the U-statistic U_n provides an unbiased estimator of $\theta(F)$. Furthermore, U_n possesses the optimality property of being the best unbiased estimator (minimum-variance unbiased estimator) whenever $\mathcal{P}$ is sufficiently large, for instance, if it includes all distributions for which $\theta(F)$ is finite (see, e.g., Serfling, 1980, Section 5.1.4).

Some examples of U-statistics are presented below.

Example 1. *The simplest case is obtained for kernels of degree one. Taking $\Phi(x) = x$, the corresponding U-statistic is*

$$U_n = \frac{1}{n}\sum_{i=1}^{n} \Phi(X_i) = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}_n.$$

Hence, the sample mean is obtained as a U-statistic with kernel $\Phi(x) = x$.

Example 2. *The Gini mean difference admits a representation as a U-statistic of degree two. Namely,*

$$U_n = \binom{n}{2}^{-1} \sum_{1\leq i<j\leq n} \Phi(X_i, X_j),$$

where the symmetric kernel is given by

$$\Phi(x, y) = |x - y|.$$

Example 3. *Let $X_1, \ldots, X_n$ be continuous observations and let R_i be the rank assigned to $|X_i|$ in the ordered sequence of absolute observations $|X_1|, \ldots, |X_n|$. The Wilcoxon signed-rank statistic is defined as*

$$W_n^+ = \sum_{i=1}^{n} R_i \mathbf{1}\{X_i > 0\}.$$

It represents the sum of the ranks corresponding to positive observations. This statistic can be expressed as a linear combination of two one-sample U-statistics. Indeed,

$$W_n^+ = \binom{n}{2} U_n^{(1)}(h) + n U_n^{(2)}(f),$$

where

$$U_n^{(1)}(h) = \binom{n}{2}^{-1} \sum_{1 \le i < j \le n} h(X_i, X_j), \qquad U_n^{(2)}(f) = \frac{1}{n} \sum_{i=1}^{n} f(X_i),$$

with kernels

$$h(x, y) = \mathbf{1}\{x + y > 0\}, \qquad f(x) = \mathbf{1}\{x > 0\}.$$

After introducing U-statistics and several illustrative examples, the next object of interest is the class of V-statistics, which form a closely related class of statistics. In contrast to U-statistics, where the summation is taken over increasing tuples of indices and therefore only over mutually distinct observations, V-statistics are defined by averaging the kernel over all possible m-tuples of indices. Thus, diagonal terms, corresponding to cases in which two or more arguments of the kernel are evaluated at the same observation, are also included. Accordingly, for a kernel Φ of degree m, the corresponding V-statistic is defined by

$$V_n = V_n(\Phi) = \frac{1}{n^m} \sum_{i_1=1}^{n} \cdots \sum_{i_m=1}^{n} \Phi(X_{i_1}, \ldots, X_{i_m}), \tag{1.1}$$

where $\{X_{i_1}, \ldots, X_{i_m}\}$ are m elements chosen from $\{X_1, \ldots, X_n\}$ with replacement, and the summation runs over all n^m ordered m-tuples.

Example 4. *(Rizzo, 2002) Let $X_1, \ldots, X_n$ denote a random sample from a distribution F, and suppose that the null hypothesis is $H_0 : F = F_0$. The one-sample energy gof statistic*

can be written as a V-statistic of degree two. Indeed, if ρ is a distance function such that all the expectations below are finite, then

$$\mathcal{E}_n = \frac{2}{n}\sum_{i=1}^{n}\mathbb{E}_Y\rho(X_i, Y) - \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\rho(X_i, X_j) - \mathbb{E}\rho(Y, Y'),$$

where Y, Y' are independent random variables with distribution F_0. Equivalently,

$$\mathcal{E}_n = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}h(X_i, X_j),$$

with

$$h(x, x') = \mathbb{E}_Y\rho(x, Y) + \mathbb{E}_Y\rho(x', Y) - \rho(x, x') - \mathbb{E}\rho(Y, Y').$$

Thus, $\mathcal{E}_n$ is a one-sample V-statistic of degree two.

Example 5. *Let $X_1, \ldots, X_n$ be a random sample from a distribution function F, and suppose that the null hypothesis is $H_0 : F = F_0$. The Cramér-von Mises statistic defined by*

$$T_n = \int_{\mathbb{R}} \big(F_n(t) - F_0(t)\big)^2 \, dF_0(t).$$

where

$$F_n(t) = \frac{1}{n}\sum_{i=1}^{n}\mathbf{1}\{X_i \le t\},$$

can be written as the V-statistic of degree two

$$T_n = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\Phi(X_i, X_j),$$

with kernel

$$\Phi(x, y) = \int_{\mathbb{R}} \big(\mathbf{1}\{x \le t\} - F_0(t)\big)\big(\mathbf{1}\{y \le t\} - F_0(t)\big) \, dF_0(t).$$

Having presented several examples, attention is now focused on the variance of U-statistics, which provides an important basis for their asymptotic analysis. The results stated below are based on the theory derived in Hoeffding, 1948.

Assume that Φ is a symmetric kernel of degree m such that

$$\mathbb{E}\big(\Phi(X_1, \ldots, X_m)\big) = \theta$$

is well defined and

$$\mathrm{Var}\big(\Phi(X_1, \ldots, X_m)\big) < \infty.$$

For $1 \leq c \leq m$, define the centered projection of order c by

$$\Phi_c(x_1, \ldots, x_c) = \mathbb{E}\big(\Phi(X_1, \ldots, X_c, X_{c+1}, \ldots, X_m) \,\big|\, X_1 = x_1, \ldots, X_c = x_c\big) - \theta.$$

Let $\{\widetilde{X}_1, \widetilde{X}_2, \ldots\}$ be an independent copy of $\{X_1, X_2, \ldots\}$, defined on the same probability space. For $1 \leq c \leq m$, set

$$\delta_c = \mathrm{Cov}\big(\Phi(X_1, \ldots, X_c, X_{c+1}, \ldots, X_m), \Phi(X_1, \ldots, X_c, \widetilde{X}_{c+1}, \ldots, \widetilde{X}_m)\big). \tag{1.2}$$

Now, it holds

$$\begin{aligned}\delta_c &= \mathbb{E}\big(\Phi(X_1, \ldots, X_c, X_{c+1}, \ldots, X_m)\, \Phi(X_1, \ldots, X_c, \tilde{X}_{c+1}, \ldots, \tilde{X}_m)\big) \\ &\quad - \big(\mathbb{E}\, \Phi(X_1, \ldots, X_m)\big)^2 \\ &= \mathbb{E}\big(\Phi_c(X_1, \ldots, X_c) + \theta\big)^2 - \theta^2 = \mathrm{Var}\big(\Phi_c(X_1, \ldots, X_c)\big) \geq 0.\end{aligned}$$

The variance of the U-statistic U_n is then derived as follows.

$$\begin{aligned}\mathrm{Var}(U_n) &= \binom{n}{m}^{-2} \sum_{1 \leq i_1 < \cdots < i_m \leq n} \sum_{1 \leq j_1 < \cdots < j_m \leq n} \mathrm{Cov}\big(\Phi(X_{i_1}, \ldots, X_{i_m}), \Phi(X_{j_1}, \ldots, X_{j_m})\big) \\ &= \binom{n}{m}^{-1} \sum_{c=1}^{m} \binom{m}{c} \binom{n-m}{m-c} \delta_c.\end{aligned}$$

Definition 1.4. *The kernel associated with the U-statistic U_n of degree m is said to be nondegenerate if $\delta_1 > 0$, and degenerate if $\delta_1 = 0$. More generally, the kernel is said to be degenerate of order $c - 1$ if*

$$\delta_1 = \cdots = \delta_{c-1} = 0 \quad \textit{and} \quad \delta_c > 0,$$

for some $2 \leq c \leq m$. The special case $c = 2$ is referred to as weak degeneracy.

Let U_n be a U-statistic whose kernel is degenerate of order $c - 1$. Then it holds:

$$\mathrm{Var}(U_n) = \frac{c!\binom{m}{c}^2 \delta_c}{n^c} + O\left(n^{-c-1}\right),\ n \to \infty. \tag{1.3}$$

The corresponding projection of the U-statistic U_n is defined as

$$\widehat{U}_n = \theta + \binom{m}{c} \binom{n}{c}^{-1} \sum_{1 \leq i_1 < \cdots < i_c \leq n} \Phi_c(X_{i_1}, \ldots, X_{i_c}).$$

Consequently, $U_n - \widehat{U}_n$ is itself a U-statistic with kernel Φ^*

$$U_n - \widehat{U}_n = \binom{n}{m}^{-1} \sum_{1 \le i_1 < \cdots < i_m \le n} \Phi^*(X_{i_1}, \ldots, X_{i_m}),$$

where

$$\Phi^*(x_1, \ldots, x_m) = \Phi(x_1, \ldots, x_m) - \theta - \sum_{1 \le i_1 < \cdots < i_c \le m} \Phi_c(x_{i_1}, \ldots, x_{i_c}) \tag{1.4}$$

By construction, the kernel of the U-statistic $U_n - \widehat{U}_n$ is degenerate of order c, i.e.

$$\delta_1^{(\Phi^*)} = \cdots = \delta_c^{(\Phi^*)} = 0.$$

Here, $\delta_i^{(\Phi^*)}$ is defined analogously to δ_i in (1.2) for $i = 1, \ldots, c$, with the kernel Φ replaced by Φ^*. Hence, using (1.3), yields

$$\boxed{\mathbb{E}\left(U_n - \widehat{U}_n\right)^2 = O\left(n^{-c-1}\right),\ n \to \infty.}$$

In contrast to U-statistics, V-statistics contain terms corresponding to repeated indices. More precisely, let

$$V_n = \frac{1}{n^m} \sum_{i_1=1}^{n} \cdots \sum_{i_m=1}^{n} \Phi(X_{i_1}, \ldots, X_{i_m})$$

be a V-statistic with symmetric kernel Φ such that $\mathbb{E}(\Phi(X_1, \ldots, X_m)) = \theta$, and let U_n denote the corresponding U-statistic. Then

$$V_n = \frac{m!\binom{n}{m}}{n^m} U_n + \frac{1}{n^m} \sum_{\substack{(i_1,\ldots,i_m) \in \{1,\ldots,n\}^m \\ |\{i_1,\ldots,i_m\}| < m}} \Phi(X_{i_1}, \ldots, X_{i_m}). \tag{1.5}$$

Assume that the kernel Φ is degenerate of order $c - 1$. Now, projection of the V-statistic V_n can be defined as

$$\widehat{V}_n = \theta + \binom{m}{c} \frac{1}{n^c} \sum_{i_1,\ldots,i_c=1}^{n} \Phi_c(X_{i_1}, \ldots, X_{i_c}).$$

Thus, the projection argument carries over directly to the part formed by mutually distinct indices. However, one must also account for the additional terms arising from repeated indices. If these terms are negligible in mean square,

namely if

$$\mathbb{E}\Big(\frac{1}{n^m}\sum_{\substack{(i_1,\ldots,i_m)\in\{1,\ldots,n\}^m\\ |\{i_1,\ldots,i_m\}|<m}}\Phi^*(X_{i_1},\ldots,X_{i_m})\Big)^2 = O(n^{-c-1}), \tag{1.6}$$

where Φ^* is the kernel defined in (1.4), then it follows:

$$\boxed{\mathbb{E}\left(V_n-\widehat{V}_n\right)^2 = O(n^{-c-1}),\ \ n\to\infty.} \tag{1.7}$$

1.2.1 Asymptotic properties of U- and V- statistics

This section summarizes the main limit results for U- and V-statistics, which will be used in the asymptotic analysis of the test statistics considered in the dissertation.

Theorem 1.6 (Strong law of large numbers for U-statistics, Hoeffding, 1961)**.** *Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables and let U_n be a U-statistic of degree m with symmetric kernel Φ. If $\mathbb{E}\left(\Phi(X_1,\ldots,X_m)\right) = \theta$ and $\mathbb{E}\left|\Phi(X_1,\ldots,X_m)\right| < \infty$, then*

$$U_n - \theta \xrightarrow[n\to\infty]{a.s.} 0.$$

The following theorem gives the corresponding limiting result in the non-degenerate case.

Theorem 1.7 (Hoeffding, 1948)**.** *Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables and let U_n be a U-statistic of degree m with symmetric kernel Φ such that $\mathbb{E}\left(\Phi(X_1,\ldots,X_m)\right) = \theta$ and $\mathbb{E}\left(\Phi^2(X_1,\ldots,X_m)\right) < \infty$. If $\delta_1 > 0$, then*

$$\sqrt{n}\big(U_n-\theta\big) \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}\left(0, m^2\delta_1\right).$$

In the weakly degenerate case, the second projection becomes the leading term in the asymptotic expansion. The corresponding limiting result is stated below.

Theorem 1.8 (Serfling, 1980)**.** *Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables with distribution F and support $S \subset \mathbb{R}^d$. Let U_n be a U-statistic of degree m with kernel Φ such that $\mathbb{E}\left(\Phi(X_1,\ldots,X_m)\right) = \theta$ and $\mathbb{E}\left(\Phi^2(X_1,\ldots,X_m)\right) < \infty$. Assume that the kernel is weakly degenerate, i.e. $\delta_2 > \delta_1 = 0$. Let $A : L^2(S,F) \to L^2(S,F)$ be the integral operator defined by*

$$Ag(y) = \int \Phi_2(x,y)g(x)\,dF(x), \tag{1.8}$$

and let $\{\lambda_k\}_{k\geq 1}$ *be its nonzero eigenvalues. Then*

$$n(U_n - \theta) \xrightarrow[n\to\infty]{\mathcal{D}} \binom{m}{2} \sum_{k=1}^{\infty} \lambda_k (Z_k^2 - 1),$$

where $\{Z_k\}_{k\geq 1}$ *are i.i.d. standard normal random variables.*

In the preceding theorem, $L^2(S, F)$ denotes the space of all real-valued measurable functions on S that are square-integrable with respect to F, that is,

$$L^2(S) = \left\{ f : S \to \mathbb{R} \;\middle|\; \int_S |f(x)|^2 \, dF(x) < \infty \right\}.$$

In the case of V-statistics, the presence of terms with repeated indices requires additional care. The decomposition in (1.5) separates the off-diagonal part of the V-statistic from the terms involving repeated indices. This reduction allows the following result to be derived from the strong law of large numbers for U-statistics stated in Theorem 1.6. A proof along these lines can be found, for example, in Korolyuk and Borovskikh, 1994.

Theorem 1.9 (Strong law of large numbers for V-statistics)**.** *Let* $X_1, X_2, \ldots, X_n$ *be i.i.d. random variables and let* V_n *be a V-statistic of degree m with symmetric kernel* Φ*. Assume that* $\mathbb{E}\Phi(X_1, \ldots, X_m) = \theta$ *and* $\mathbb{E}\left|\Phi(X_{i_1}, \ldots, X_{i_m})\right| < \infty$ *for* $1 \leq i_1, \ldots, i_m \leq m$*. Then*

$$V_n \xrightarrow[n\to\infty]{a.s.} \theta.$$

The following lemma gives a comparison result for the U- and V-statistics associated with the same kernel. It shows that, under suitable moment conditions, their difference is asymptotically negligible.

Lemma 1.2 (Serfling, 1980)**.** *Let* $X_1, X_2, \ldots, X_n$ *be i.i.d. random variables and let* U_n *and* V_n *be a U- and V-statistic of degree m with symmetric kernel* Φ*, respectively. Let r be a positive integer. Suppose that*

$$\mathbb{E}\left|\Phi\left(X_{i_1}, \ldots, X_{i_m}\right)\right|^r < \infty$$

for all choices $1 \leq i_1, \ldots, i_m \leq m$*. Then*

$$\mathbb{E}\left|U_n - V_n\right|^r = O\left(n^{-r}\right), \; n \to \infty.$$

The following theorem follows by combining the preceding lemma, applied

with $r = 2$, with the relation (1.5) and the asymptotic result for nondegenerate U-statistics presented in Theorem 1.7.

Theorem 1.10. *Let $X_1, \ldots, X_n$ be i.i.d. random variables and let V_n be a V-statistic of degree m with symmetric kernel Φ such that $\mathbb{E}\left(\Phi(X_1, \ldots, X_m)\right) = \theta$ and $\mathbb{E}\left(\Phi^2(X_{i_1}, \ldots, X_{i_m})\right) < \infty$, for all $1 \leq i_1, \ldots, i_m \leq n$. If $\delta_1 > 0$, then*

$$\sqrt{n}(V_n - \theta) \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}\left(0, m^2\delta_1\right).$$

For the weakly degenerate case, consider the V-statistic V_n

$$V_n(\Phi) = \frac{1}{n^m} \sum_{i_1,\ldots,i_m=1}^{n} \Phi(X_{i_1}, \ldots, X_{i_m}),$$

where Φ is a symmetric kernel of degree m such that $\mathbb{E}\Phi(X_1, \ldots, X_m) = \theta$ and assume $0 = \delta_1 < \delta_2$. Using the decomposition (1.5), the V-statistic can be written as

$$V_n(\Phi) - \theta = \frac{m!\binom{n}{m}}{n^m} U_n(\Phi) - \theta + D_n + R_n,$$

where $U_n(\Phi)$ is the associated U-statistics with kernel Φ,

$$D_n = \frac{\binom{m}{2}}{n^m} \sum_{\substack{i_1,\ldots,i_{m-1}=1 \\ i_1,\ldots,i_{m-1} \text{ distinct}}}^{n} \Phi(X_{i_1}, X_{i_1}, X_{i_2}, \ldots, X_{i_{m-1}})$$

is the contribution of the terms with exactly one pair of equal indices, while R_n contains all remaining terms. By the symmetry of Φ, all terms with exactly one repeated pair have the same limiting contribution. Assume, in addition, that

$$\mathbb{E}\left|\Phi(X_{i_1}, X_{i_2}, \ldots, X_{i_m})\right| < \infty \text{ for all } 1 \leq i_1, \ldots, i_m \leq m. \tag{1.9}$$

Then, by the strong law of large numbers for U-statistics (Theorem 1.6), it follows that

$$nD_n \xrightarrow[n\to\infty]{a.s.} \binom{m}{2} \mathbb{E}\Phi(X_1, X_1, X_2, \ldots, X_{m-1}). \tag{1.10}$$

The terms contained in R_n involve at most $m - 2$ distinct indices and therefore, under the imposed moment assumption (1.9),

$$nR_n = o_{\mathbb{P}}(1). \tag{1.11}$$

By the definition of the centered second projection,

$$\Phi_2(x, y) = \mathbb{E}\left[\Phi(x, y, X_3, \ldots, X_m)\right] - \theta.$$

Hence,

$$\mathbb{E}\Phi_2(X_1, X_1) = \mathbb{E}\Phi(X_1, X_1, X_2, \ldots, X_{m-1}) - \theta.$$

Since Φ is weakly degenerate, limit theorem for degenerate U-statistics (Theorem 1.8) yields

$$n(U_n(\Phi) - \theta) \xrightarrow[n\to\infty]{\mathcal{D}} \binom{m}{2} \sum_{k=1}^{\infty} v_k(Z_k^2 - 1), \tag{1.12}$$

where $(Z_k)_{k\geq 1}$ are independent standard normal random variables and $(v_k)_{k\geq 1}$ are the eigenvalues of the integral operator generated by Φ_2. The relations (1.12), (1.10), and (1.11), together with Slutsky's theorem (Theorem 1.3) and the continuous mapping theorem (Theorem 1.2), imply the following result. The proof for the case $m = 2$ can be found, for example, in Serfling, 1980, Section 6.4.

Theorem 1.11. *Let $X_1, \ldots, X_n$ be i.i.d. random variables with distribution F and support $S \subset \mathbb{R}^d$. Let V_n be a V-statistic of degree m with symmetric kernel Φ such that $\mathbb{E}\left(\Phi(X_1, \ldots, X_m)\right) = 0$, $\mathbb{E}\left(\Phi^2(X_1, \ldots, X_m)\right) < \infty$ and $\mathbb{E}|\Phi(X_{i_1}, \ldots, X_{i_m})| < \infty$ for all $1 \leq i_1, \ldots, i_m \leq m$. If $\delta_2 > \delta_1 = 0$, then*

$$nV_n \xrightarrow[n\to\infty]{\mathcal{D}} \binom{m}{2} \left[\sum_{k=1}^{\infty} v_k(Z_k^2 - 1) + \mathbb{E}\Phi_2(X, X)\right],$$

where $Z_1, Z_2, \ldots$ are independent standard normal random variables, and $v_1, v_2, \ldots$ are the eigenvalues of the integral operator A defined in (1.8).

Remark 1.1. *In particular, assume that the integral operator A is positive semidefinite, trace class and that the trace formula*

$$\operatorname{tr}(A) = \int_S \Phi_2(x, x)\, dF(x) \tag{1.13}$$

holds. Then the limiting distribution takes the form

$$n(V_n - \theta) \xrightarrow[n\to\infty]{\mathcal{D}} \binom{m}{2} \sum_{k=1}^{\infty} v_k Z_k^2.$$

Consequently, the variance of the limiting distribution of $n(V_n - \theta)$ is given by

$$2\binom{m}{2}^2 \sum_{k=1}^{\infty} v_k^2 = 2\binom{m}{2}^2 \iint_{S\times S} \Phi_2^2(x,y)\, dF(x) dF(y). \tag{1.14}$$

1.2.2 Asymptotic properties of a linear combination of two degenerate V-statistics

This subsection provides novel theoretical results for the distribution of finite linear combinations of certain weakly degenerate V-statistics, addressing both cases with known parameters and cases in which parameters are estimated from the data.

The asymptotic properties of linear combinations of degenerate (including weakly degenerate) U-statistics with specific weighting schemes have been studied in Yamato et al., 2005. In contrast, the present analysis is concerned with two related cases. The first involves linear combinations of weakly degenerate V-statistics with arbitrary kernels, while the second concerns V-statistics with estimated parameters, under additional regularity conditions on the kernels that are commonly satisfied in practice (see, e.g., Meintanis, 2004; Meintanis et al., 2007; Meintanis et al., 2015; Ejsmont et al., 2023; Ndwandwe et al., 2023; Meintanis et al., 2024a; Meintanis et al., 2024b).

Let $X_1, \ldots, X_n$ be an i.i.d. sample from a distribution function F with support $S \subset \mathbb{R}^d$. Consider a V-statistic $V_n(\lambda)$ whose kernel depends on an unknown parameter $\lambda = (\lambda_1, \ldots, \lambda_r) \in \Lambda \subset \mathbb{R}^r$. If λ is replaced by a consistent estimator $\hat{\lambda}_n$, the statistic $V_n(\hat{\lambda}_n)$ is referred to as a V-statistic with estimated parameters.

Assume that $V_n(\lambda)$ and $W_n(\lambda)$ are zero-mean weakly degenerate V-statistics, with the following representation:

$$\begin{aligned} V_n(\lambda) &= \frac{1}{n^{m_1}} \sum_{i_1=1}^{n} \cdots \sum_{i_{m_1}=1}^{n} \Phi(X_{i_1}, \ldots, X_{i_{m_1}}; \lambda), \\ W_n(\lambda) &= \frac{1}{n^{m_2}} \sum_{i_1=1}^{n} \cdots \sum_{i_{m_2}=1}^{n} \Psi(X_{i_1}, \ldots, X_{i_{m_2}}; \lambda), \end{aligned} \tag{1.15}$$

where $\lambda = (\lambda_1, \ldots, \lambda_r)$ is an unknown r-dimensional parameter. Here, $\Phi(\cdot;\lambda)$ and $\Psi(\cdot;\lambda)$ represent the corresponding symmetric weakly degenerate kernels with zero expectation. The following theorem presents the limiting distribution of linear combinations of $V_n(\hat{\lambda}_n)$ and $W_n(\hat{\lambda}_n)$, where $\hat{\lambda}_n$ is a consistent estimator of the true parameter λ.

Theorem 1.12. *Let $V_n(\lambda)$ and $W_n(\lambda)$ be weakly degenerate V-statistics for every $\lambda \in \Lambda \subset \mathbb{R}^r$, with symmetric zero-mean kernels $\Phi(\cdot;\lambda)$ and $\Psi(\cdot;\lambda)$ of degrees $m_1 \leq m_2$, respectively, defined on the same i.i.d. sample $X_1, \ldots, X_n$. Let $V_n(\hat{\lambda}_n)$ and $W_n(\hat{\lambda}_n)$ denote the corresponding V-statistics obtained by replacing λ with its estimator $\hat{\lambda}_n$. Furthermore, assume that*

$$\mathbb{E}\Phi^2(X_1, \ldots, X_{m_1}; \lambda) < \infty, \qquad \mathbb{E}\Psi^2(X_1, \ldots, X_{m_2}; \lambda) < \infty,$$

for all $\lambda \in \Lambda$. For $p = (p_1, p_2)$, where $p_1, p_2 \geq 0$ and $p_1 + p_2 > 0$, define

$$L_n(\lambda) = p_1 V_n(\lambda) + p_2 W_n(\lambda).$$

Suppose that the following conditions hold:

(i) $\sqrt{n}(\hat{\lambda}_n - \lambda_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \alpha(X_i) + o_{\mathbb{P}}(1)$ *for some $\lambda_0 = (\lambda_{0,1}, \ldots, \lambda_{0,r}) \in \Lambda$, where $\alpha(x) = (\alpha_1(x), \ldots, \alpha_r(x))$ has zero mean with a finite second moment;*

(ii) the functions $\Phi(\cdot;t)$ and $\Psi(\cdot;t)$ are three times continuously differentiable with respect to t on Γ, and there exist positive random variables $Y_\Phi^{(d)}$ and $Y_\Psi^{(d)}$, for $d = 1, 2, 3$, such that

(a)

$$\sup_{\gamma \in \Gamma} \sum_{i=1}^{r} \left| \frac{\partial \Phi}{\partial t_i}\Big|_{t=\gamma} \right| < Y_\Phi^{(1)} \quad \text{and} \quad \sup_{\gamma \in \Gamma} \sum_{i=1}^{r} \left| \frac{\partial \Psi}{\partial t_i}\Big|_{t=\gamma} \right| < Y_\Psi^{(1)};$$

(b)

$$\sup_{\gamma \in \Gamma} \sum_{i,j=1}^{r} \left| \frac{\partial^2 \Phi}{\partial t_i \partial t_j}\Big|_{t=\gamma} \right| < Y_\Phi^{(2)} \quad \text{and} \quad \sup_{\gamma \in \Gamma} \sum_{i,j=1}^{r} \left| \frac{\partial^2 \Psi}{\partial t_i \partial t_j}\Big|_{t=\gamma} \right| < Y_\Psi^{(2)};$$

(c)

$$\sup_{\gamma \in \Gamma} \sum_{i,j,k=1}^{r} \left| \frac{\partial^3 \Phi}{\partial t_i \partial t_j \partial t_k}\Big|_{t=\gamma} \right| \leq Y_\Phi^{(3)} \quad \text{and} \quad \sup_{\gamma \in \Gamma} \sum_{i,j,k=1}^{r} \left| \frac{\partial^3 \Psi}{\partial t_i \partial t_j \partial t_k}\Big|_{t=\gamma} \right| \leq Y_\Psi^{(3)},$$

where $\mathbb{E}(Y_\Phi^{(d)}) < \infty$ and $\mathbb{E}(Y_\Psi^{(d)}) < \infty$ for $d = 1, 2, 3$.

If, in addition, the operator A^* generated by

$$\xi_2^*(x,y;\lambda_0,p) = p_1\frac{m_1(m_1-1)}{m_2(m_2-1)}\phi_2(x,y;\lambda_0) + p_2\psi_2(x,y;\lambda_0) + \frac{1}{m_2(m_2-1)}\sum_{j_1,j_2=1}^{r}\alpha_{j_1}(x)\alpha_{j_2}(y)\left(p_1\mathbb{E}\left[\frac{\partial^2\Phi}{\partial\lambda_{j_1}\partial\lambda_{j_2}}\right] + p_2\mathbb{E}\left[\frac{\partial^2\Psi}{\partial\lambda_{j_1}\partial\lambda_{j_2}}\right]\right), \tag{1.16}$$

where $x, y \in S$, is positive semidefinite, trace class and the diagonal trace formula

$$\mathrm{tr}(A^*) = \int_S \xi_2^*(x,x;\lambda_0,p)\,dF(x)$$

holds, then all corresponding eigenvalues ν_j, $j \geq 1$, are nonnegative, and

$$nL_n(\hat{\lambda}_n) \xrightarrow[n\to\infty]{\mathcal{D}} \binom{m_2}{2}\sum_{j=1}^{\infty}\nu_j Z_j^2,$$

where $Z_1, Z_2, \ldots$ are independent standard normal random variables.

Proof. Using a Taylor expansion of $L_n(\hat{\lambda}_n)$ around the true parameter value λ_0 yields

$$nL_n(\hat{\lambda}_n) = nL_n(\lambda_0) + n\sum_{j=1}^{r}\frac{\partial L_n(\lambda_0)}{\partial\lambda_j}(\hat{\lambda}_{n,j}-\lambda_{0,j}) + \frac{n}{2}\sum_{j_1,j_2=1}^{r}\frac{\partial^2 L_n(\lambda_0)}{\partial\lambda_{j_1}\partial\lambda_{j_2}}(\hat{\lambda}_{n,j_1}-\lambda_{0,j_1})(\hat{\lambda}_{n,j_2}-\lambda_{0,j_2}) + nR_n, \tag{1.17}$$

where R_n is the Taylor remainder.

The statistic $V_n(\lambda)$, which is originally of degree m_1, can be written in an equivalent form as a V-statistic of degree m_2. Hence, the linear combination $L_n(\lambda)$ can also be treated as a V-statistic of degree m_2. Since its kernel is degenerate for λ in a neighbourhood Γ of λ_0, differentiation of the degeneracy condition with respect to λ_j shows that the kernel corresponding to $\partial L_n(\lambda_0)/\partial\lambda_j$ is also degenerate. The interchange of differentiation and expectation is justified by condition $(ii)(a)$. Therefore,

$$\frac{\partial L_n(\lambda_0)}{\partial\lambda_j} = O_{\mathbb{P}}(n^{-1}), \quad j = 1,\ldots,r.$$

In addition, using condition (i),

$$\hat{\lambda}_{n,j} - \lambda_{0,j} = O_{\mathbb{P}}(1/\sqrt{n}),\ j = 1,...,r,$$

hence the second term on the right-hand side of (1.17) is $o_{\mathbb{P}}(1)$. Condition $(ii)(c)$ further implies that $nR_n = o_{\mathbb{P}}(1)$.

Under condition $(ii)(b)$, the law of large numbers for V-statistics implies that $\partial^2 L_n(\lambda_0)/(\partial\lambda_{j_1}\partial\lambda_{j_2})$ converges to its expectation. By combining the preceding relation with the condition (i), it follows that

$$nL_n(\hat{\lambda}_n) = nL_n^*(\lambda_0) + o_P(1)$$

where

$$\begin{aligned} L_n^*(\lambda_0) &= L_n(\lambda_0) + \frac{1}{2}\sum_{j_1,j_2=1}^{r}\Big(p_1\mathbb{E}\Big[\frac{\partial^2\Phi}{\partial\lambda_{j_1}\partial\lambda_{j_2}}\Big] + p_2\mathbb{E}\Big[\frac{\partial^2\Psi}{\partial\lambda_{j_1}\partial\lambda_{j_2}}\Big]\Big)\frac{1}{n^2}\sum_{i_1,i_2=1}^{n}\alpha_{j_1}(X_{i_1})\alpha_{j_2}(X_{i_2}) \\ &= L_n(\lambda_0) + \frac{1}{2n^2}\sum_{i_1,i_2=1}^{n}\sum_{j_1,j_2=1}^{r}\alpha_{j_1}(X_{i_1})\alpha_{j_2}(X_{i_2})\Big(p_1\mathbb{E}\Big[\frac{\partial^2\Phi}{\partial\lambda_{j_1}\partial\lambda_{j_2}}\Big] + p_2\mathbb{E}\Big[\frac{\partial^2\Psi}{\partial\lambda_{j_1}\partial\lambda_{j_2}}\Big]\Big) \end{aligned}$$

Notice that $nL_n^*(\lambda_0)$ can be expressed as a V-statistic of degree m_2 with an appropriately symmetrized kernel whose first projection is equal to zero, while the second projection is given in (1.16).

□

Remark 1.2. *It is clear that the generalization of Theorem 1.12 to the case of a finite linear combination of weakly degenerate V-statistics is straightforward.*

Corollary 1.1. *Let V_n and W_n be two weakly degenerate V-statistics with symmetric zero-mean kernels Φ and Ψ of degrees $m_1 \leq m_2$, respectively, defined on the same sample $X_1, \ldots, X_n$. Assume also that $\mathbb{E}\Phi^2(X_1, X_2, ..., X_{m_1}) < \infty$ and $\mathbb{E}\Psi^2(X_1, X_2, ..., X_{m_2}) < \infty$. For $p = (p_1, p_2)$, with $p_1, p_2 \geq 0$ and $p_1 + p_2 > 0$, set*

$$L_n = p_1 V_n + p_2 W_n.$$

If the operator A generated by

$$\xi_2(x, y; p) = p_1\frac{m_1(m_1-1)}{m_2(m_2-1)}\phi_2(x, y) + p_2\psi_2(x, y),\ x, y \in S$$

is positive semidefinite, trace class and the diagonal trace formula

$$\mathrm{tr}(A) = \int_S \xi_2(x, x; p)\, dF(x)$$

holds, then the corresponding eigenvalues ν_j, $j \geq 1$, are nonnegative and

$$nL_n \xrightarrow[n\to\infty]{\mathcal{D}} \binom{m_2}{2} \sum_{j=1}^{\infty} \nu_j Z_j^2.$$

Corollary 1.2. *Let V_n and $W_n(\lambda)$, $\lambda \in \Lambda \subset \mathbb{R}^r$, be weakly degenerate V-statistics defined on the same i.i.d. sample $X_1, \ldots, X_n$, with symmetric zero-mean kernels Φ and $\Psi(\cdot;\lambda)$ of degrees $m_1 \leq m_2$, respectively. Let $W_n(\hat{\lambda}_n)$ denote the V-statistic corresponding to $W_n(\lambda)$, obtained by replacing λ with its estimator $\hat{\lambda}_n$. Assume also that*

$$\mathbb{E}\Phi^2(X_1, \ldots, X_{m_1}) < \infty, \qquad \mathbb{E}\Psi^2(X_1, \ldots, X_{m_2}; \lambda) < \infty,$$

for all $\lambda \in \Lambda \subset \mathbb{R}^r$. Then, for $p = (p_1, p_2)$ such that $p_1, p_2 \geq 0$ and $p_1 + p_2 > 0$ set $L_n(\lambda) = p_1 V_n + p_2 W_n(\lambda)$. Assume that the following conditions hold:

(i) $\sqrt{n}(\hat{\lambda}_n - \lambda_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \alpha(X_i) + o_{\mathbb{P}}(1)$, for some $\lambda_0 = (\lambda_{0,1}, \ldots, \lambda_{0,r}) \in \Lambda$, where $\alpha(x) = (\alpha_1(x), \ldots, \alpha_r(x))$ has zero mean with a finite second moment;

(ii) the function $\Psi(\cdot; t)$ is three times continuously differentiable with respect to t on Γ, and, for each $d = 1, 2, 3$, there exists a positive random variable $Y^{(d)}$ such that $\mathbb{E}Y^{(d)} < \infty$ and

(a)
$$\sup_{\gamma\in\Gamma} \sum_{i=1}^r \left| \frac{\partial \Psi}{\partial t_i}\Big|_{t=\gamma} \right| < Y^{(1)},$$

(b)
$$\sup_{\gamma\in\Gamma} \sum_{i,j=1}^r \left| \frac{\partial^2 \Psi}{\partial t_i \partial t_j}\Big|_{t=\gamma} \right| < Y^{(2)},$$

(c)
$$\sup_{\gamma\in\Gamma} \sum_{i,j,k=1}^r \left| \frac{\partial^3 \Psi}{\partial t_i \partial t_j \partial t_k}\Big|_{t=\gamma} \right| \leq Y^{(3)}.$$

If, in addition, the operator A^* generated by

$$\begin{aligned} \xi_2^*(x, y; \lambda_0, p) = {} & p_1 \frac{m_1(m_1-1)}{m_2(m_2-1)} \phi_2(x, y) + p_2 \psi_2(x, y; \lambda_0) \\ & + \frac{1}{m_2(m_2-1)} \sum_{j_1, j_2=1}^r \alpha_{j_1}(x) \alpha_{j_2}(y) p_2 \mathbb{E}\left[\frac{\partial^2 \Psi}{\partial \lambda_{j_1} \partial \lambda_{j_2}} \right], \quad x, y \in S. \end{aligned} \tag{1.18}$$

is positive semidefinite and trace class, and the diagonal trace formula

$$\operatorname{tr}(A^*) = \int_S \xi_2^*(x, x; \lambda_0, p)\, dF(x)$$

holds, then the corresponding eigenvalues ν_j, $j \geq 1$, are nonnegative and

$$nL_n(\hat{\lambda}_n) \xrightarrow[n\to\infty]{\mathcal{D}} \binom{m_2}{2} \sum_{j=1}^{\infty} \nu_j Z_j^2.$$

1.3 U-processes

The theory introduced in the preceding section concerns U-statistics generated by a single symmetric kernel. In some applications, however, a family of kernels indexed by a class of sets must be considered, giving rise to a corresponding collection of U-statistics. After centering and normalization, the resulting stochastic process is referred to as a U-process. Only the particular form needed later in the dissertation is presented in this section; see Schneemeier, 1989 for a more general treatment.

Let $X_1, X_2, \ldots$ be i.i.d. random variables with common distribution F and support $S \subseteq \mathbb{R}^d$. Fix integers $m, r \geq 1$, and let

$$\psi_k : S^m \longrightarrow \mathbb{R}, \quad k = 1, \ldots, r,$$

be measurable functions. For $t_1, \ldots, t_r \in \mathbb{R}$, define

$$C_{t_1,\ldots,t_r} = \{(x_1, \ldots, x_m) \in S^m : \psi_1(x_1, \ldots, x_m) \leq t_1, \ldots, \psi_r(x_1, \ldots, x_m) \leq t_r\},$$

and consider the indexing class

$$\mathcal{C} = \{C_{t_1,\ldots,t_r} : (t_1, \ldots, t_r) \in \mathbb{R}^r\}.$$

For every $C \in \mathcal{C}$, let $H(C) = \mathbb{P}\{(X_1, \ldots, X_m) \in C\}$. The empirical counterpart of $H(C)$, for $n \geq m$, is defined by

$$H_n(C) = \frac{1}{n(n-1)\cdots(n-m+1)} \sum_{\substack{1 \leq i_1, \ldots, i_m \leq n \\ i_j \neq i_k,\ 1 \leq j < k \leq m}} \mathbf{1}\{(X_{i_1}, \ldots, X_{i_m}) \in C\}, \quad C \in \mathcal{C}.$$

The functions $\psi_1, \ldots, \psi_r$ and consequently the sets in $\mathcal{C}$ need not be symmetric with respect to their arguments. Nevertheless, $H_n(C)$ can be represented as a

U-statistic with a symmetric kernel. Indeed, let Π_m denote the set of all permutations of $\{1,\dots,m\}$ and, for $C \in \mathcal{C}$, define

$$h_C(x_1,\dots,x_m) = \frac{1}{m!}\sum_{\pi\in\Pi_m} \mathbf{1}\big\{(x_{\pi(1)},\dots,x_{\pi(m)}) \in C\big\}.$$

Then h_C is symmetric and

$$H_n(C) = \binom{n}{m}^{-1} \sum_{1\le i_1<\cdots<i_m\le n} h_C(X_{i_1},\dots,X_{i_m}).$$

Since $X_1,\dots,X_m$ are i.i.d., their joint distribution is invariant under permutations. Therefore,

$$\mathbb{E}\big(h_C(X_1,\dots,X_m)\big) = H(C),$$

and consequently $H_n(C)$ is an unbiased estimator of $H(C)$.

Definition 1.5. *For each $n \ge m$, the stochastic process*

$$\mathbb{U}_n = \big\{\mathbb{U}_n(C) : C \in \mathcal{C}\big\},$$

defined by

$$\mathbb{U}_n(C) = \sqrt{n}\big\{H_n(C) - H(C)\big\}, \quad C \in \mathcal{C},$$

is called the U-process indexed by $\mathcal{C}$.

The supremum norm of $\mathbb{U}_n$ is defined by

$$\|\mathbb{U}_n\|_{\mathcal{C}} = \sup_{C\in\mathcal{C}} |\mathbb{U}_n(C)|.$$

For $C = C_{t_1,\dots,t_r}$, it is convenient to write

$$H(t_1,\dots,t_r) = H(C_{t_1,\dots,t_r})$$

and

$$H_n(t_1,\dots,t_r) = H_n(C_{t_1,\dots,t_r}).$$

More explicitly,

$$H(t_1,\dots,t_r) = \mathbb{P}\left\{\psi_1(X_1,\dots,X_m) \le t_1, \dots \psi_r(X_1,\dots,X_m) \le t_r\right\},$$

whereas

$$H_n(t_1,\dots,t_r) = \frac{1}{n(n-1)\cdots(n-m+1)} \sum_{\substack{1\le i_1,\dots,i_m\le n \\ i_j\neq i_l,\ 1\le j<l\le m}} \prod_{k=1}^{r} \mathbf{1}\left\{\psi_k(X_{i_1},\dots,X_{i_m}) \le t_k\right\}.$$

Thus, H is the joint distribution function of the random variables $\psi_1(X_1,\dots,X_m),\dots,\psi_r(X_1,\dots,X_m)$, while H_n is its joint U-empirical counterpart. Correspondingly, the U-process indexed by $(t_1,\dots,t_r) \in \mathbb{R}^r$ can be written as

$$\mathbb{U}_n(t_1,\dots,t_r) = \sqrt{n}\big\{H_n(t_1,\dots,t_r) - H(t_1,\dots,t_r)\big\}.$$

Example 6. *In the special case $d = m = r = 1$ and $\psi_1(x) = x$, the class $\mathcal{C}$ reduces to*

$$\mathcal{C} = \big\{(-\infty, t] \cap S : t \in \mathbb{R}\big\}.$$

Moreover,

$$H(t) = \mathbb{P}\{X_1 \le t\} = F(t) \quad \textit{and} \quad H_n(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{X_i \le t\} = F_n(t),$$

so that

$$\mathbb{U}_n(t) = \sqrt{n}\{F_n(t) - F(t)\}.$$

Hence, the ordinary empirical process is a special case of a U-process.

The following result gives the almost sure bound for the supremum norm. An application of Theorem 3.7 of Schneemeier, 1989 requires the indexing class to be a Vapnik-Chervonenkis class of sets. The indexing class $\mathcal{C}$ satisfies this condition, since the class of lower orthants in $\mathbb{R}^r$ is a Vapnik-Chervonenkis class and $\mathcal{C}$ is its inverse-image class under the fixed mapping

$$\Psi = (\psi_1,\dots,\psi_r) : S^m \longrightarrow \mathbb{R}^r.$$

Therefore, Theorem 3.7 of Schneemeier, 1989 yields the following result.

Theorem 1.13. *Let $(a_n)_{n\in\mathbb{N}} \subset (1,\infty)$ be a sequence such that*

$$\sum_{n=1}^{\infty} a_n^{-\varepsilon} < \infty \qquad \textit{for every } \varepsilon > 0.$$

Then

$$(\log a_n)^{-1/2} \|\mathbb{U}_n\|_{\mathcal{C}} \xrightarrow[n\to\infty]{\text{a.s.}} 0.$$

Taking $a_n = e^n$ in this theorem, it follows that

$$\frac{1}{\sqrt{n}}\|\mathbb{U}_n\|_{\mathcal{C}} \xrightarrow[n\to\infty]{\text{a.s.}} 0.$$

By the definitions of $\mathbb{U}_n$ and $\mathcal{C}$, this is equivalent to

$$\sup_{(t_1,\ldots,t_r)\in\mathbb{R}^r} |H_n(t_1,\ldots,t_r) - H(t_1,\ldots,t_r)| \xrightarrow[n\to\infty]{\text{a.s.}} 0.$$

For $m = r = 2$, this establishes the uniform convergence result used in Section 3.1.

1.4 Characterizations of probability distributions

As already mentioned, characterization theorems play an important role in the study of probability distributions and stochastic structures, since they provide necessary and sufficient conditions under which such objects are uniquely determined. They make it possible to study distributions through structural, functional, or probabilistic properties rather than only through explicit analytical forms.

On the other hand, characterizations also provide a natural basis for gof testing and model validation. Depending on the underlying property, characterization-based tests may rely on independence relations, equality in distribution, regression-type identities, properties of order statistics or records, moment and reliability properties, or functional identities involving characteristic functions, Laplace transforms, probability generating functions, and related transforms. In this dissertation, particular emphasis is placed on independence-type characterizations, whose general form is presented below.

Let $X_1, \ldots, X_{\max(n,m)}$ be i.i.d. random variables with distribution function F, and let $\psi_1 : \mathbb{R}^n \to \mathbb{R}$ and $\psi_2 : \mathbb{R}^m \to \mathbb{R}$ be two measurable sample functionals. Then the following equivalence holds:

$$\psi_1(X_1, \ldots, X_n) \text{ and } \psi_2(X_1, \ldots, X_m) \text{ are independent}$$

iff the distribution function F belongs to a specified family $\mathcal{F}$.

In the sequel, several independence-type characterizations are presented, which will be used in the construction of the gof tests later in the dissertation.

Example 7 (Ferguson, 1967)**.** *Let X and Y be independent, discrete, non-degenerate*

random variables. Then

$$U = \min(X, Y) \qquad \textit{and} \qquad W = |X - Y|$$

are independent iff the distributions of X and Y may, by a simultaneous change of location and scale, be put into one of the following four forms.

(i) For some $0 < r_1 < 1$ and $0 < r_2 < 1$,

$$\mathbb{P}\{X = k\} = (1 - r_1) r_1^k, \qquad k = 0, 1, \ldots,$$

and

$$\mathbb{P}\{Y = k\} = (1 - r_2) r_2^k, \qquad k = 0, 1, \ldots.$$

(ii) For some integer $n \geq 1$, $0 < r < 1$, and $-1 \leq \theta \leq 1$,

$$\mathbb{P}\{X = k\} = \frac{(1 - r) r^k}{1 + \theta r^n} \begin{cases} 1, & k = 0, 1, \ldots, n - 1, \\ 1 + \theta, & k = n, n + 1, \ldots, \end{cases}$$

and

$$\mathbb{P}\{Y = k\} = \frac{(1 - r) r^k}{1 - \theta r^n} \begin{cases} 1, & k = 0, 1, \ldots, n - 1, \\ 1 - \theta, & k = n, n + 1, \ldots. \end{cases}$$

(iii) For some $0 < r < 1$ and $0 < \theta < \infty$, either

$$\mathbb{P}\{X = k\} = \frac{(1 - r) r^k}{1 + \theta r^2} \begin{cases} 1, & k = 0, 1, \\ 1 + \theta, & k = 2, 3, \ldots, \end{cases}$$

and

$$\mathbb{P}\{Y = k\} = \begin{cases} (1 + \theta r)^{-1}, & k = 0, \\ 1 - (1 + \theta r)^{-1}, & k = 1, \end{cases}$$

or the same with X and Y interchanged.

(iv) For some $0 < r < 1$ and $0 < \theta \leq \infty$,

$$\mathbb{P}\{X = k\} = (1 - r^2) r^k \begin{cases} (1 + \theta r)^{-1}, & k = 0, 2, 4, \ldots, \\ \theta (1 + \theta r)^{-1}, & k = 1, 3, 5, \ldots, \end{cases}$$

and

$$\mathbb{P}\{Y=k\} = (1-r^2)r^k \begin{cases} r\theta(\theta+r)^{-1}, & k=0,2,4,\ldots, \\ (\theta+r)^{-1}, & k=1,3,5,\ldots. \end{cases}$$

Here, a simultaneous change of location and scale for the pair (X,Y) corresponds to the transformation

$$(X,Y) \mapsto (a+bX,\ a+bY),$$

for real a and $b>0$.

Example 8 (Lukacs, 1955)**.** *Let X and Y be i.i.d. random variables with strictly positive, absolutely continuous distribution F. Then, $\psi_1(X,Y) = \frac{X}{X+Y}$ and $\psi_2(X,Y) = X+Y$ are independent iff F belongs to the family of gamma $\Gamma(\alpha,\beta)$ distributions with density*

$$f(x;\alpha,\beta) = \frac{1}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \beta^{\alpha}, \quad x>0, \quad \alpha>0, \quad \beta>0.$$

Example 9 (Khatri, 1962)**.** *Let the X and Y be i.i.d. random variables with distribution function F and such that $E(X)$, $E(X^2)$, $E(\frac{1}{X})$ and $E(\frac{1}{X+Y})$ are non-null and finite. Then $\psi_1(X,Y) = X+Y$ and $\psi_2(X,Y) = 1/X + 1/Y - 4/(X+Y)$ are independent iff F belongs to te family of inverse Gaussian $\mathrm{IG}(\mu,\lambda)$ distributions with density*

$$f(x;\mu,\lambda) = \sqrt{\frac{\lambda}{2\pi}} x^{-3/2} \exp\left(-\frac{\lambda(x-\mu)^2}{2\mu^2 x}\right), \quad x>0, \quad \mu>0, \quad \lambda>0.$$

Example 10 (Arnold, 1979)**.** *Let X,Y be i.i.d. random variables with a continuous distribution. Then $\psi_1(X,Y) = X$ and $\psi_2(X,Y) = \frac{X+Y}{1-XY}$ are independent iff X has the standard Cauchy $\mathcal{C}(0,1)$ distribution with density*

$$f(x) = \frac{1}{\pi(1+x^2)}, \quad x \in \mathbb{R}.$$

Chapter 2

Goodness-of-fit tests for the geometric distribution based on Ferguson's characterization

Classical edf-based tests constitute a standard approach to goodness-of-fit testing. Although these tests can be applied to count data, their null distributions generally depend on the particular discrete model under consideration. This has motivated the development of alternative procedures, including tests based on the difference between the empirical pgf and the theoretical one (see, e.g., Rueda and O'Reilly 1999), as well as tests constructed from the empirical counterpart of the differential equation satisfied by the pgf (see, e.g., Batsidis et al. 2020, Milošević et al. 2021). Alongside these approaches, tests based on characteristic functions have also been proposed (see, e.g., Jiménez-Gamero et al. 2009).

The present chapter is devoted to the construction of gof tests for the geometric distribution based on an independence-type characterization. In the context of count data, tests relying on characterizations of this kind are still rather limited; to the best of our knowledge, the test proposed in Jiménez-Gamero and Alba-Fernández, 2021 is the only closely related contribution. This motivates the investigation of independence-type characterizations as a tool for developing gof procedures for discrete distributions.

2.1 Proposed test statistics

For $p \in (0,1)$, let $\mathcal{G}(p)$ denote the geometric distribution with support $\mathbb{N} = \{1,2,\ldots\}$ and pmf

$$\mathbb{P}\{X = k\} = (1-p)^{k-1}p,\ k \in \mathbb{N}, \tag{2.1}$$

where p is the probability of success. Let X be a random variable taking values in $\mathbb{N} = \{1, 2, \ldots\}$, with unknown distribution, and let $X_1, \ldots, X_n$ denote i.i.d. observations from the distribution of X. Based on these observations, the objective is to construct novel gof tests for the composite null hypothesis

$$H_0 : X \sim \mathcal{G}(p), \text{ for some } p \in (0, 1),$$

versus the alternative:

$$H_1 : X \not\sim \mathcal{G}(p), \text{ for all } p \in (0, 1).$$

The tests are constructed based on the Ferguson characterization given in Example 7. Under the additional assumption that X and Y are identically distributed, this characterization reduces to the form stated below.

Characterization 2.1. *Let X and Y be nondegenerate, discrete i.i.d. random variables. Then $\min\{X, Y\}$ and $|X - Y|$ are independent iff X and Y have geometric distribution up to a simultaneous location and scale transformation.*

Remark 2.1. *In Ferguson, 1967, the geometric distribution is introduced in a more general form, with location and scale parameters α and $\beta > 0$. More precisely, a random variable X is said to have a geometric distribution if its pmf is of the form*

$$\mathbb{P}\{X = k\} = (1 - p)p^{\frac{k-\alpha}{\beta}}, \quad k \in \{\alpha, \alpha + \beta, \alpha + 2\beta, \ldots\}.$$

Thus, the support of X is not necessarily $\{1, 2, \ldots\}$, but may be shifted and rescaled.

If X has such a distribution, then there exist real constants a and $b > 0$ such that the transformed variable $a + bX$ has the probability law given in (2.1)*. Consequently, a simultaneous change of location and scale for the pair (X, Y) corresponds to the transformation*

$$(X, Y) \mapsto (a + bX,\ a + bY).$$

In other words, there exist real constants a and $b > 0$ such that both $a + bX$ and $a + bY$ follow the distribution in (2.1)*. In particular, when the support of X and Y is $\{1, 2, \ldots\}$, no such transformation is needed, and the independence of $\min(X, Y)$ and $|X - Y|$ characterizes the probability law in* (2.1) *uniquely.*

Suppose that X and Y are nondegenerate discrete i.i.d. random variables supported on $\{1, 2, \ldots\}$. Motivated by the above characterization, consider the joint

pgf of the random vector $\big(\min\{X,Y\}, |X-Y|\big)$, defined by

$$g(t,s) = \mathbb{E}\left(t^{\min\{X,Y\}}s^{|X-Y|}\right),\ (t,s) \in [0,1]^2.$$

The condition

$$g(t,s) = g(t,1)g(1,s),\ \text{for all}\ (t,s) \in [0,1]^2,$$

holds iff X and Y have the geometric distribution given by (2.1) for some p. Although pgf depends on the parameter p under the geometric distribution, this dependence is suppressed in the notation for simplicity. Let $X_1, X_2, ..., X_n$ be an i.i.d. sample with the same common distribution as X and Y. Then

$$g_n(t,s) = \frac{1}{n^2}\sum_{i,j=1}^{n} t^{\min\{X_i,X_j\}}s^{|X_i-X_j|}$$

is the joint V-empirical pgf of the random vector $(\min\{X,Y\}, |X-Y|)$. Therefore, it is meaningful to consider

$$D_n(t,s) = g_n(t,s) - g_n(t,1)g_n(1,s)$$

as the primary component for the construction of a gof test, since under H_0, $D_n(t,s)$ is expected to be small. Values significantly different from zero may indicate a departure from the geometric distribution.

Before proceeding with the proposal, let $L^2([0,1]^2)$ denote the separable Hilbert space of measurable functions $f : [0,1]^2 \to \mathbb{R}$ satisfying

$$\int_0^1\int_0^1 f^{\,2}(t,s)dtds < \infty.$$

The inner product and the norm on this space are defined, respectively, by $\langle f,g\rangle_{L^2} = \int_0^1\int_0^1 f(t,s)g(t,s)dtds$ and $\|f\|_{L^2} = \sqrt{\langle f,f\rangle_{L^2}}$. At this point, two novel test statistics can be introduced:

$$I_n = \langle D_n, 1\rangle_{L^2} = \int_0^1\int_0^1 D_n(t,s)dtds, \tag{2.2}$$

and

$$T_n = ||D_n||^2_{L^2} = \int_0^1 \int_0^1 D_n^2(t,s)dtds. \tag{2.3}$$

After integration, the resulting expressions for I_n and T_n are as follows:

$$\begin{aligned}
I_n =& \frac{1}{n^2} \sum_{i,j=1}^{n} \frac{1}{\min(X_i, X_j) + 1} \frac{1}{|X_i - X_j| + 1} \\
&- \frac{1}{n^4} \sum_{i,j=1}^{n} \frac{1}{\min(X_i, X_j) + 1} \sum_{i,j=1}^{n} \frac{1}{|X_i - X_j| + 1}, \\
T_n =& \frac{1}{n^4} \sum_{i,j,k,l=1}^{n} \frac{1}{\min(X_i, X_j) + \min(X_k, X_l) + 1} \frac{1}{|X_i - X_j| + |X_k - X_l| + 1} \\
&- \frac{2}{n^6} \sum_{i,j,k,l,r,s=1}^{n} \frac{1}{\min(X_i, X_j) + \min(X_k, X_l) + 1} \frac{1}{|X_i - X_j| + |X_r - X_s| + 1} \\
&+ \frac{1}{n^8} \sum_{i,j,k,l=1}^{n} \frac{1}{\min(X_i, X_j) + \min(X_k, X_l) + 1} \sum_{i,j,k,l=1}^{n} \frac{1}{|X_i - X_j| + |X_k - X_l| + 1}.
\end{aligned}$$

If the null hypothesis is true, then the values of these test statistics should be close to zero. The statistic I_n provides a computationally less demanding option. Since it can take both positive and negative values, evidence against the null hypothesis is associated with large absolute values of I_n. I_n is not a true distance, since it may remain small even when $D_n(t,s)$ deviates substantially from zero. However, such behavior is rarely observed in practice (see, e.g., Cuparić et al. 2019). This may be explained by the fact that, under the alternatives considered, the integrand tends to have a predominantly positive or predominantly negative sign. In contrast, T_n is based on a squared discrepancy, so large values of T_n indicate evidence against the null hypothesis.

The following section is devoted to the asymptotic analysis of the proposed test statistics. The almost sure limiting behavior is first established. Subsequently, auxiliary lemmas are presented and used to derive the corresponding weak convergence results.

2.2 Asymptotic properties

The aim of this section is to derive the stochastic limits of the proposed test statistics and establish their limiting distributions under the null hypothesis.

Theorem 2.1 (Halaj and Milošević, 2025)**.** *Suppose that $X_1, ..., X_n$ are i.i.d. random variables with the same distribution as a random variable X taking values in $\mathbb{N}$. Then*

$$I_n \xrightarrow[n\to\infty]{a.s.} I = \int_0^1 \int_0^1 \big(g(t,s) - g(t,1)g(1,s)\big)dtds,$$

and

$$T_n \xrightarrow[n\to\infty]{a.s.} T = \int_0^1 \int_0^1 \big(g(t,s) - g(t,1)g(1,s)\big)^2 dtds,$$

where $g(t,s), g(t,1)$ and $g(1,s)$ are the joint and marginal pgfs of the random vector $(\min\{X_1, X_2\}, |X_1 - X_2|)$, respectively.

Proof. From Proposition 1 in Novoa-Muñoz and Jiménez-Gamero, 2014, it follows that

$$\sup_{(t,s)\in[0,1]^2} |g_n(t,s) - g(t,s)| \xrightarrow[n\to\infty]{a.s.} 0.$$

An analogous convergence result holds for the differences between the marginal empirical and marginal theoretical pgfs. Therefore, if

$$D(t,s) = g(t,s) - g(t,1)g(1,s)$$

denotes the theoretical quantity and $D_n(t,s)$ stands for its empirical counterpart, then it follows

$$\begin{aligned}
|D_n(t,s) - D(t,s)| =& |g_n(t,s) - g_n(t,1)g_n(1,s) - g(t,s) + g(t,1)g(1,s)| \\
=& |g_n(t,s) - g(t,s) - g_n(t,1)(g_n(1,s) - g(1,s) + g(1,s)) \\
&+ g(1,s)(g(t,1) - g_n(t,1) + g_n(t,1))| \\
\leq& |g_n(t,s) - g(t,s)| + |g_n(t,1)(g_n(1,s) - g(1,s))| \\
&+ |g(1,s)(g(t,1) - g_n(t,1))| \\
\leq& |g_n(t,s) - g(t,s)| + |g_n(1,s) - g(1,s)| \\
&+ |g(t,1) - g_n(t,1)|.
\end{aligned}$$

So, the following result holds

$$\sup_{(t,s)\in[0,1]^2} |D_n(t,s) - D(t,s)| \xrightarrow[n\to\infty]{a.s.} 0.$$

Since I_n and T_n are continuous functions of $D_n(t,s)$, their $a.s.$ convergence follows

directly from the continuous mapping theorem. □

Corollary 2.1. *Suppose that $X \sim \mathcal{G}(p)$, for some $p \in (0,1)$, and let $X_1, \ldots, X_n$ be independent copies of X. Then*

$$I_n \xrightarrow[n\to\infty]{a.s.} 0 \quad \textit{and} \quad T_n \xrightarrow[n\to\infty]{a.s.} 0.$$

Theorem 2.1 indicates that, since $T = 0$ iff the integrand is equal to zero, T_n is consistent against any fixed alternative that shares the same support as (2.1). Establishing suitable critical regions requires knowledge of the test statistic's null distribution or its approximation. As the exact null distributions of T_n and I_n are unknown, their asymptotic null distributions are derived in the following. Several preliminary lemmas are introduced first.

Lemma 2.1 (Halaj and Milošević, 2025)**.** *Suppose that $X \sim \mathcal{G}(p)$, for some $p \in (0,1)$, and let $X_1, \ldots, X_n$ be independent copies of X. Then, it holds*

$$\sqrt{n}\big(g_n(t,s) - g(t,1)g(1,s)\big) = \frac{2}{\sqrt{n}} \sum_{i=1}^{n} \big(\xi(X_i;t,s,p) - g(t,1)g(1,s)\big) + r_n(t,s),$$

for $(t,s) \in [0,1]^2$, where $\xi(x;t,s,p)$ is defined as

$$\xi(x;t,s,p) = \begin{cases} \frac{ps(t(1-p))^x}{1-s(1-p)} + \frac{pt(s^x-(t(1-p))^x)}{s-t(1-p)}, & s \neq t(1-p), \\ t^x(1-p)^{x-1}p(x-1) - \frac{t^x(1-p)^{x-1}p}{(1-p)^2t-1}, & s = t(1-p). \end{cases} \tag{2.4}$$

and $||r_n||_{L^2} = o_{\mathbb{P}}(1)$ as $n \to \infty$.

Proof. Let

$$g_n(t,s) = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \Phi(X_i, X_j; t, s), \quad (t,s) \in [0,1]^2,$$

where

$$\Phi(X_i, X_j; t, s) = t^{\min\{X_i, X_j\}} s^{|X_i - X_j|}.$$

Then $g_n(t,s)$ is a V-statistic of degree 2 whose first projection is

$$\xi(X_1;t,s,p) = \mathbb{E}\big(\Phi(X_1, X_2; t, s) \mid X_1\big).$$

Now,

$$\xi(x;t,s,p) = \sum_{y=1}^{\infty} t^{\min\{x,y\}} s^{|x-y|} p(1-p)^{y-1}$$

$$= \sum_{y=1}^{x} t^y s^{x-y} p(1-p)^{y-1} + \sum_{y=x+1}^{\infty} t^x s^{y-x} p(1-p)^{y-1}.$$

Therefore, for $s \neq t(1-p)$, applying the formula for a finite geometric sum to the first sum and the formula for a convergent infinite geometric series to the second yields the following

$$\boxed{\xi(x;t,s,p) = pt\frac{s^x - ((1-p)t)^x}{s-(1-p)t} + \frac{p(1-p)^x t^x s}{1-(1-p)s}.}$$

In the special case $s = t(1-p)$, the first sum is evaluated directly, which gives

$$\boxed{\xi(x;t,s,p) = px(1-p)^{x-1}t^x + \frac{p(1-p)^{x+1}t^{x+1}}{1-(1-p)^2 t},}$$

thereby obtaining the expression given in (2.4). By the Hoeffding decomposition, it follows that

$$\sqrt{n}\big(g_n(t,s) - g(t,1)g(1,s)\big) = \frac{2}{\sqrt{n}}\sum_{i=1}^{n}\big(\xi(X_i;t,s,p) - g(t,1)g(1,s)\big) + r_n(t,s).$$

Here,

$$r_n(t,s) = \sqrt{n}\left(g_n(t,s) - \widehat{g}_n(t,s)\right),$$

where

$$\widehat{g}_n(t,s) = g(t,1)g(1,s) + \frac{2}{n}\sum_{i=1}^{n}\left\{\xi(X_i;t,s,p) - g(t,1)g(1,s)\right\}.$$

The difference $g_n(t,s) - \widehat{g}_n(t,s)$ is a V-statistic with kernel

$$\Phi^*(x,y;t,s) = \Phi(x,y;t,s) - \xi(x;t,s,p) - \xi(y;t,s,p) + g(t,1)g(1,s).$$

whose first projection is identically equal to zero. Moreover, all $O(\cdot)$-terms below are understood uniformly with respect to $(t,s) \in [0,1]^2$. Condition (1.6) reduces to

$$\mathbb{E}\left(\frac{1}{n^2}\sum_{i=1}^{n}\Phi^*(X_i,X_i;t,s)\right)^2 = O(n^{-2}). \tag{2.5}$$

Using the fact that

$$\begin{aligned} 0 \leq \xi(x;t,s,p) &= \mathbb{E}(\Phi(x,X_2;t,s)) \\ &= \mathbb{E}(t^{\min(x,X_2)}s^{|x-X_2|}) \leq \sum_{k=1}^{\infty} p(1-p)^{k-1} = 1 \end{aligned}$$

for all $x \in \mathbb{N}$ and $(t,s) \in [0,1]^2$, it follows that $\Phi^*(x,y;t,s)$ is uniformly bounded for all $x,y \in \mathbb{N}$ and $(t,s) \in [0,1]^2$, and the required condition is therefore fulfilled. Hence, by (1.7),

$$\mathbb{E}\left(g_n(t,s) - \widehat{g}_n(t,s)\right)^2 = O(n^{-2}).$$

and it follows

$$\mathbb{E}r_n^2(t,s) = O(n^{-1}).$$

Now, by Tonelli's theorem,

$$\mathbb{E}\|r_n\|_{L^2}^2 = \mathbb{E}\int_0^1\int_0^1 r_n^2(t,s)\,dt\,ds = \int_0^1\int_0^1 \mathbb{E}r_n^2(t,s)\,dt\,ds.$$

Using Markov's inequality, for every $\varepsilon > 0$,

$$P\left(\|r_n\|_{L^2} > \varepsilon\right) \leq \frac{\mathbb{E}\|r_n\|_{L^2}^2}{\varepsilon^2} \to 0, \quad \text{for } n \to \infty.$$

Hence, $\|r_n\|_{L^2} = o_{\mathbb{P}}(1)$. □

Lemma 2.2 (Halaj and Milošević, 2025)**.** *Suppose that $X \sim \mathcal{G}(p)$, for some $p \in (0,1)$, and let $X_1,\ldots,X_n$ be independent copies of X. Then it holds*

$$\sqrt{n}\left(g_n(t,s) - g_n(t,1)g_n(1,s)\right) = \frac{2}{\sqrt{n}}\sum_{i=1}^{n}\eta(X_i;t,s,p) + \rho_n(t,s), \quad (t,s) \in [0,1]^2$$

where

$$\eta(x;t,s,p) = \xi(x;t,s,p) - g(1,s)\xi(x;t,1,p) - g(t,1)\xi(x;1,s,p) + g(t,1)g(1,s),$$

with $\xi(x;t,s,p)$ defined in (2.4) and $||\rho_n||_{L^2} = o_{\mathbb{P}}(1)$.

Proof. Let

$$\begin{aligned} g_n(t,s) - g_n(t,1)g_n(1,s) = {} & g_n(t,s) - g(t,1)g(1,s) + g(t,1)g(1,s) \\ & - \left(g_n(t,1) - g(t,1) + g(t,1)\right)\left(g_n(1,s) - g(1,s) + g(1,s)\right) \end{aligned}$$

$$
\begin{aligned}
&= g_n(t,s) - g(t,1)g(1,s) - g(1,s)\big(g_n(t,1) - g(t,1)\big) \\
&- g(t,1)\big(g_n(1,s) - g(1,s)\big) - \epsilon_n(t,s),
\end{aligned}
$$

where $\epsilon_n(t,s) = \big(g_n(t,1) - g(t,1)\big)\big(g_n(1,s) - g(1,s)\big)$. From the previous lemma and the central limit theorem in Hilbert spaces, it follows that $||\sqrt{n}\epsilon_n||_{L^2} = o_{\mathbb{P}}(1)$. Thus, by applying the previous lemma once again and using the fact that $g(1,1) = 1$, the following is obtained

$$
\begin{aligned}
&\sqrt{n}\big(g_n(t,s) - g(t,1)g(1,s) - g(1,s)(g_n(t,1) - g(t,1)) - g(t,1)(g_n(1,s) - g(1,s))\big) \\
&\quad = \frac{2}{\sqrt{n}} \sum_{i=1}^{n} \eta(X_i; t, s, p) + \rho_n(t,s),
\end{aligned}
$$

with $||\rho_n||_{L^2} = o_{\mathbb{P}}(1)$. Now, the result follows. □

Theorem 2.2 (Halaj and Milošević, 2025)**.** *Let $X \sim \mathcal{G}(p)$, for some $p \in (0,1)$, and let $X_1, \ldots, X_n$ be independent copies of X. Then*

$$
\sqrt{n} I_n \xrightarrow[n\to\infty]{\mathcal{D}} \int_0^1 \int_0^1 Z(t,s)\, dtds \quad \text{and} \quad nT_n \xrightarrow[n\to\infty]{\mathcal{D}} \int_0^1 \int_0^1 Z^2(t,s) dtds,
$$

where $\{Z(t,s),\ (t,s) \in [0,1]^2\}$ is a centered Gaussian process in $L^2([0,1]^2)$ with covariance function

$$
c((t,s),(u,v);p) = 4\mathbb{E}\big(\eta(X;t,s,p)\eta(X;u,v,p)\big), \quad t,s,u,v \in [0,1], \tag{2.6}
$$

where

$$
\eta(x;t,s,p) = \xi(x;t,s,p) - g(1,s)\xi(x;t,1,p) - g(t,1)\xi(x;1,s,p) + g(t,1)g(1,s),
$$

with $\xi(x;t,s,p)$ defined in (2.4)*.*

Proof. Under the null hypothesis, for each $(t,s) \in [0,1]^2$,

$$
\mathbb{E}\big(\xi(X;t,s,p)\big) = g(t,s;p).
$$

As already noted, the dependence of g on the parameter p is suppressed for notational simplicity, so $g(t,s)$ will be used instead of $g(t,s;p)$. Thus,

$$
\begin{aligned}
\mathbb{E}\big(\eta(X;t,s,p)\big) = \mathbb{E}\big(\xi(X;t,s,p)\big) &- g(1,s)\mathbb{E}\big(\xi(X;t,1,p)\big) - g(t,1)\mathbb{E}\big(\xi(X;1,s,p)\big) \\
&+ g(t,1)g(1,s)
\end{aligned}
$$

$$= g(t,s) - g(1,s)g(t,1) - g(t,1)g(1,s) + g(t,1)g(1,s)$$
$$= g(t,s) - g(t,1)g(1,s) = 0.$$

To establish that $\eta(X;t,s,p)$ has finite variance for every $(t,s) \in [0,1]^2$, it is enough to show that the functions appearing in its definition are uniformly bounded.
As shown in the proof of Lemma 2.1

$$0 \leq \xi(x;t,s,p) \leq 1,$$

for all $x \in \mathbb{N}$ and $t,s \in [0,1]$.
Together with the bound $0 \leq g(t,s) \leq 1$ for all $(t,s) \in [0,1]^2$, this yields

$$|\eta(x;t,s,p)| \leq 2, \text{ for all } x \in \mathbb{N} \text{ and } t,s \in [0,1]. \tag{2.7}$$

Thus

$$\mathbb{E}[\eta(X;t,s,p)^2] < \infty \text{ for all } t,s \in [0,1].$$

Now, the proof follows from the previous lemma, the central limit theorem in Hilbert spaces and the continuous mapping theorem. □

Corollary 2.2. *The limiting distribution of $\sqrt{n}I_n$ is zero-mean normal with variance*

$$\sigma^2(p) = 4\mathbb{E}\left(\varphi_I(X;p)^2\right) < \infty, \tag{2.8}$$

where

$$\varphi_I(x;p) = \int_0^1 \int_0^1 \eta(x;t,s,p)\, dt\, ds.$$

Furthermore, the limiting distribution of nT_n admits the representation

$$\sum_{k\geq 1} \lambda_k \chi^2_{1,k},$$

where $\{\chi^2_{1,k}\}_{k\geq 1}$ is a sequence of independent chi-square random variables with one degree of freedom, and $\{\lambda_k\}_{k\geq 1}$ denotes the sequence of positive eigenvalues of the integral operator $A : L^2([0,1]^2) \to L^2([0,1]^2)$, defined by

$$Af(x,y) = \int_0^1 \int_0^1 c((x,y),(u,v);p)\, f(u,v)\, du\, dv.$$

2.3 Bootstrap approximation of the null distribution

The results in the previous section show that the limiting null distributions of the proposed test statistics are not distribution-free. Since the limiting null distributions depend on the unknown parameter p, a parametric bootstrap procedure is used to approximate the corresponding critical values and p-values. The procedure is presented in Algorithm 1, while its theoretical justification is provided in what follows.

Algorithm 1 Parametric bootstrap algorithm

1: Compute the test statistic $\mathcal{S}_{n,\text{obs}}$ based on the observed sample.

2: For a sufficiently large integer B, perform the following steps for each $b \in \{1, \ldots, B\}$:

   (a) Generate a bootstrap sample $X_1^*, X_2^*, \ldots, X_n^*$, where $X_i^* \sim \mathcal{G}(\hat{p}_n)$ and $\hat{p}_n$ denotes an estimator of p based on the observed sample;

   (b) Calculate the test statistic $\mathcal{S}_{n,b}^*$ based on the bootstrap sample $X_1^*, \ldots, X_n^*$.

3: Approximate the p-value by calculating:

$$\hat{p}_{value} = \frac{1}{B} \sum_{b=1}^{B} I\left\{\mathcal{S}_{n,b}^* > \mathcal{S}_{n,\text{obs}}\right\},$$

or, alternatively, approximate the critical value by finding the a-th order statistic of the bootstrap test statistics, denoted by $\mathcal{S}_{a:B}^*$, where $a = \lfloor (1-\alpha)B \rfloor + 1$.

.

Let $X_1, \ldots, X_n$ represent the observed sample. Conditionally on this sample, let $X_1^*, \ldots, X_n^*$ be independent bootstrap observations drawn from the geometric distribution $\mathcal{G}(\hat{p}_n)$, where $\hat{p}_n = \hat{p}_n(X_1, ..., X_n)$. Let $\mathcal{S}_n$ stand for any of the proposed test statistics. Define the bootstrap statistic $\mathcal{S}_n^*$ in the same way as $\mathcal{S}_n$, but with the original observations $X_1, \ldots, X_n$ replaced by their bootstrap counterparts $X_1^*, \ldots, X_n^*$.

The parametric bootstrap approximates the null distribution of $\mathcal{S}_n$ by the conditional distribution of the bootstrap statistic $\mathcal{S}_n^*$, namely $\mathbb{P}_*(\mathcal{S}_n^* \leq t)$, where $\mathbb{P}_*$ denotes probability given the observed sample.

The asymptotic results presented in Theorem 1.5 provide the basis for deriving the limiting behavior of the bootstrap versions of I_n and T_n. The result is stated in the following theorem.

Theorem 2.3. *Suppose that $X_1, ..., X_n$ are i.i.d. random variables with the same distribution as a random variable X taking values in $\mathbb{N}$ such that $\mathbb{E}(X) < \infty$. Assume that*

$\hat{p}_n \xrightarrow[n\to\infty]{a.s.} p_0$, *where* p_0 *is the true parameter value if the null hypothesis is true. Then, conditionally on the* $X_1, ..., X_n$,

$$\sqrt{n} I_n^* \xrightarrow[n\to\infty]{\mathcal{D}} \int_0^1 \int_0^1 Z(t,s)\, dtds \quad \textit{and} \quad nT_n^* \xrightarrow[n\to\infty]{\mathcal{D}} \int_0^1 \int_0^1 Z^2(t,s) dtds,$$

where $\{Z(t,s),\ (t,s) \in [0,1]^2\}$ *is a centered Gaussian process in* $L^2([0,1]^2)$ *with covariance function given in* (2.6).

Proof. Given the data $X_1, \ldots, X_n$, let $X_1^*, \ldots, X_n^*$ be the bootstrap sample, so that, conditionally on $X_1, \ldots, X_n$,

$$X_1^*, \ldots, X_n^* \overset{\text{i.i.d.}}{\sim} \mathcal{G}(\hat{p}_n).$$

For $x \in \mathbb{N}$ and $p \in (0,1)$, define

$$\eta_p(x)(t,s) := \eta(x;t,s,p), \quad (t,s) \in [0,1]^2,$$

where the expression for $\eta(x;t,s,p)$ is given in the statement of Theorem 2.2. Set

$$Y_{ni} := \frac{2}{\sqrt{n}} \eta_{\hat{p}_n}(X_i^*), \quad i = 1, \ldots, n.$$

It follows from (2.7) that, for each $i = 1, \ldots, n$,

$$\int_0^1 \int_0^1 |Y_{ni}(t,s)|^2\, dt\, ds < \infty \quad \mathbb{P}_*\text{- a.s.}$$

Consequently, conditionally on the data, $Y_{n1}, \ldots, Y_{nn}$ are $L^2([0,1]^2)$-valued random elements $\mathbb{P}_*$-*a.s.* Since $X_1^*, \ldots, X_n^*$ are conditionally independent and each Y_{ni} is a measurable function of X_i^*, the random elements $Y_{n1}, \ldots, Y_{nn}$ are conditionally independent.

Denote by $g(t,s;p)$ the pgf of the random vector $\big(\min\{X,Y\}, |X-Y|\big)$, where X and Y are independent random variables with common geometric distribution $\mathcal{G}(p)$, $p \in (0,1)$.
Conditionally on the data, $X_i^* \sim \mathcal{G}(\hat{p}_n)$, $i = 1, \ldots, n$, and therefore

$$\mathbb{E}_*\big(\xi(X_i^*;t,s,\hat{p}_n)\big) = g(t,s;\hat{p}_n), \quad (t,s) \in [0,1]^2.$$

Consequently, the definition of η yields

$$\mathbb{E}_*\big(\eta(X_i^*;t,s,\hat{p}_n)\big)=0,\quad (t,s)\in[0,1]^2.$$

Let $\{e_k\}_{k\geq 1}$ be an arbitrary orthonormal basis of $L^2([0,1]^2)$. By Fubini's theorem, for every $k\geq 1$ and $i=1,...,n$

$$\mathbb{E}_*\big(\langle Y_{ni},e_k\rangle\big)=\frac{2}{\sqrt{n}}\int_0^1\int_0^1\mathbb{E}_*\big(\eta_{\hat{p}_n}(X_i^*)(t,s)\big)e_k(t,s)\,dt\,ds=0.$$

On the other hand, by (2.7),

$$\mathbb{E}_*\|Y_{ni}\|_{L^2}^2<\infty,\ i=1,\dots,n.$$

Let

$$S_n:=\sum_{i=1}^n Y_{ni}=\frac{2}{\sqrt{n}}\sum_{i=1}^n\eta_{\hat{p}_n}(X_i^*),$$

and let Γ_n denote the covariance operator of S_n. The next step is to verify conditions *(i)-(iii)* of Theorem 1.5.

(i) Fix $k,\ell\geq 1$. Since, conditionally on the data, the random elements $Y_{n1},\dots,Y_{nn}$ are independent and centered, the cross terms vanish. Hence

$$\langle\Gamma_n e_k,e_\ell\rangle=\mathbb{E}_*\big(\langle S_n,e_k\rangle\langle S_n,e_\ell\rangle\big)=4\mathbb{E}_*\big(\langle\eta_{\hat{p}_n}(X_1^*),e_k\rangle\langle\eta_{\hat{p}_n}(X_1^*),e_\ell\rangle\big).$$

Define

$$h_{k\ell}(x,p):=\langle\eta_p(x),e_k\rangle\langle\eta_p(x),e_\ell\rangle,\quad x\in\mathbb{N},\ \ p\in(0,1).$$

For fixed $x\in\mathbb{N}$ and $(t,s)\in[0,1]^2$, the mapping $p\mapsto\eta(x;t,s,p)$ is continuous on $(0,1)$. Hence, since $\hat{p}_n\xrightarrow[n\to\infty]{\mathbb{P}\text{-}a.s.}p_0$, the continuous mapping theorem yields

$$\eta(x;t,s,\hat{p}_n)\xrightarrow[n\to\infty]{\mathbb{P}\text{-}a.s.}\eta(x;t,s,p_0).$$

Moreover, for each fixed $x\in\mathbb{N}$, using (2.7), the dominated convergence theorem yields

$$\|\eta_{\hat{p}_n}(x)-\eta_{p_0}(x)\|_{L^2}\xrightarrow[n\to\infty]{\mathbb{P}\text{-}a.s.}0.$$

Hence, by the continuity of the inner product in $L^2[0,1]^2$,

$$h_{k\ell}(x,\hat{p}_n)\xrightarrow[n\to\infty]{\mathbb{P}\text{-}a.s.}h_{k\ell}(x,p_0).$$

Since, conditionally on data, $X_1^* \sim \mathcal{G}(\hat{p}_n)$, it follows that

$$\langle \Gamma_n e_k, e_\ell \rangle = 4 \sum_{x=1}^{\infty} h_{k\ell}(x, \hat{p}_n) \hat{p}_n (1 - \hat{p}_n)^{x-1}.$$

Fix ω such that $\hat{p}_n(\omega) \to p_0$, as $n \to \infty$. Then, for all sufficiently large n,

$$\hat{p}_n(\omega) \geq \frac{p_0}{2}, \qquad 1 - \hat{p}_n(\omega) \leq 1 - \frac{p_0}{2} =: \rho < 1.$$

Using the Cauchy-Schwarz inequality together with the relation (2.7) yields

$$|h_{k\ell}(x, p)| \leq \|\eta_p(x)\|_{L^2}^2 \leq 4, \ \forall x \in \mathbb{N}, \ \ p \in (0, 1).$$

Consequently, for all sufficiently large n,

$$\left| h_{k\ell}(x, \hat{p}_n) \hat{p}_n (1 - \hat{p}_n)^{x-1} \right| \leq 4 \rho^{x-1}, \ x \in \mathbb{N},$$

and

$$\sum_{x=1}^{\infty} 4\rho^{x-1} < \infty.$$

Therefore, by the dominated convergence theorem for series,

$$\langle \Gamma_n e_k, e_\ell \rangle \xrightarrow[n\to\infty]{\mathbb{P}\text{-}a.s.} 4 \sum_{x=1}^{\infty} h_{k\ell}(x, p_0) p_0 (1 - p_0)^{x-1} = 4\mathbb{E}\big(\langle \eta_{p_0}(X), e_k \rangle \langle \eta_{p_0}(X), e_\ell \rangle \big) =: \alpha_{k\ell},$$

where $X \sim \mathcal{G}(p_0)$.

(ii) By the definition of Γ_n and the fact that summands below are nonnegative, Tonelli's theorem and Parseval's identity yield

$$\begin{aligned} \sum_{k=1}^{\infty} \langle \Gamma_n e_k, e_k \rangle &= 4 \sum_{k=1}^{\infty} \mathbb{E}_* \big(\langle \eta_{\hat{p}_n}(X_1^*), e_k \rangle^2 \big) \\ &= 4\mathbb{E}_* \left(\sum_{k=1}^{\infty} \langle \eta_{\hat{p}_n}(X_1^*), e_k \rangle^2 \right) \\ &= 4\mathbb{E}_* \|\eta_{\hat{p}_n}(X_1^*)\|_{L^2}^2. \end{aligned}$$

As in the proof of *(i)*, for each fixed $x \geq 1$,

$$\|\eta_{\hat{p}_n}(x)\|_{L^2}^2 \xrightarrow[n\to\infty]{\mathbb{P}\text{-}a.s.} \|\eta_{p_0}(x)\|_{L^2}^2 \quad \text{and} \quad 0 \leq \|\eta_{\hat{p}_n}(x)\|_{L^2}^2 \leq 4.$$

Since, conditionally on the data, $X_1^* \sim \mathcal{G}(\hat{p}_n)$, it follows

$$\mathbb{E}_* \|\eta_{\hat{p}_n}(X_1^*)\|_{L^2}^2 = \sum_{x=1}^{\infty} \|\eta_{\hat{p}_n}(x)\|_{L^2}^2 \hat{p}_n (1-\hat{p}_n)^{x-1}.$$

Then, for all sufficiently large n,

$$\left| \|\eta_{\hat{p}_n}(x)\|_{L^2}^2 \hat{p}_n (1-\hat{p}_n)^{x-1} \right| \leq 4\rho^{x-1}, \ x \in \mathbb{N},$$

where

$$\sum_{x=1}^{\infty} 4\rho^{x-1} < \infty.$$

Therefore, dominated convergence for series yields

$$\mathbb{E}_* \|\eta_{\hat{p}_n}(X_1^*)\|_{L^2}^2 \xrightarrow[n\to\infty]{\mathbb{P}\text{-}a.s.} \sum_{x=1}^{\infty} \|\eta_{p_0}(x)\|_{L^2}^2 p_0 (1-p_0)^{x-1} = \mathbb{E}\|\eta_{p_0}(X)\|_{L^2}^2.$$

On the other hand, again by Tonelli and Parseval,

$$\sum_{k=1}^{\infty} \alpha_{kk} = 4 \sum_{k=1}^{\infty} \mathbb{E}\big(\langle \eta_{p_0}(X), e_k \rangle^2\big) = 4\mathbb{E}\big(\sum_{k=1}^{\infty} \langle \eta_{p_0}(X), e_k \rangle^2\big) = 4\mathbb{E}\|\eta_{p_0}(X)\|_{L^2}^2.$$

Hence,

$$\sum_{k=1}^{\infty} \langle \Gamma_n e_k, e_k \rangle \xrightarrow[n\to\infty]{\mathbb{P}\text{-}a.s.} \sum_{k=1}^{\infty} \alpha_{kk} < \infty.$$

(iii) Let $\varepsilon > 0$ and $k \geq 1$ be fixed. Then

$$\begin{aligned} L_n(\varepsilon, e_k) &= \sum_{i=1}^{n} \mathbb{E}_* \Big(\langle Y_{ni}, e_k \rangle^2 \mathbf{1}\{ |\langle Y_{ni}, e_k \rangle| > \varepsilon \} \Big) \\ &= \sum_{i=1}^{n} \mathbb{E}_* \left(\frac{4}{n} \langle \eta_{\hat{p}_n}(X_i^*), e_k \rangle^2 \mathbf{1} \left\{ |\langle \eta_{\hat{p}_n}(X_i^*), e_k \rangle| > \frac{\varepsilon\sqrt{n}}{2} \right\} \right). \end{aligned}$$

Since, conditionally on the data, the bootstrap variables are i.i.d.,

$$L_n(\varepsilon, e_k) = 4\mathbb{E}_* \left(\langle \eta_{\hat{p}_n}(X_1^*), e_k \rangle^2 \mathbf{1} \left\{ |\langle \eta_{\hat{p}_n}(X_1^*), e_k \rangle| > \frac{\varepsilon\sqrt{n}}{2} \right\} \right).$$

By the Cauchy-Schwarz inequality,

$$|\langle \eta_{\hat{p}_n}(X_1^*), e_k \rangle| \leq \|\eta_{\hat{p}_n}(X_1^*)\|_{L^2} \, \|e_k\|_{L^2} \leq 2.$$

Therefore, whenever $\varepsilon\sqrt{n} > 4$,

$$\mathbf{1}\left\{|\langle \eta_{\hat{p}_n}(X_1^*), e_k\rangle| > \frac{\varepsilon\sqrt{n}}{2}\right\} = 0.$$

Since $\varepsilon > 0$ is fixed and $\varepsilon\sqrt{n} \to \infty$, it follows that, for all sufficiently large n,

$$L_n(\varepsilon, e_k) = 0.$$

Consequently,

$$L_n(\varepsilon, e_k) \xrightarrow[n\to\infty]{\mathbb{P}\text{-}a.s.} 0.$$

Therefore, all assumptions of Theorem 1.5 are satisfied *P-a.s.* with respect to the original sample. Now, the conclusion follows using this theorem together with the continuous mapping theorem. □

The preceding result justifies using a parametric bootstrap procedure to approximate the null distributions of the proposed test statistics. Based on the resulting bootstrap distributions, approximate critical values and p-values are computed for $|I_n|$ and T_n. In the case of I_n, the absolute value is taken because both large positive and large negative values indicate departures from the null hypothesis, whereas for T_n large values alone are relevant.

Another approach, suitable for $\sqrt{n}I_n$ in view of its asymptotic normality, is to consider the standardized statistic

$$\widetilde{I}_n = \frac{\sqrt{n}I_n}{\hat{\sigma}(\hat{p}_n)}, \tag{2.9}$$

where $\hat{p}_n = \bar{X}_n^{-1}$ is the maximum likelihood estimator of p, and $\sigma(\hat{p}_n)$ is a plug-in estimator of $\sigma(p)$, with $\sigma(p)$ given in (2.8).

Table 2.1 reports the empirical quantiles of $\widetilde{I}_n$ and the corresponding quantiles of the limiting distribution for several choices of n and p. The values were obtained by a Monte Carlo procedure with $N = 10000$ replications (see also Figure 2.1). These results show that for small sample sizes, the use of asymptotic quantiles is not justified, whereas for $n > 500$, the quantiles can be reasonably well approximated.

2.4 Power study

This section presents a simulation study designed to assess the finite-sample performance of the proposed tests. As already noted, their powers are approximated

	$n = 100$			$n = 500$			$n = 2000$			Limitting
β / p	0.2	0.5	0.8	0.2	0.5	0.8	0.2	0.5	0.8	distribution
0.99	2.70	2.83	3.20	2.60	2.62	2.72	2.58	2.64	2.60	2.58
0.95	2.07	2.07	2.29	2.00	1.99	2.01	1.96	1.99	1.95	1.96
0.9	1.75	1.72	1.66	1.69	1.67	1.65	1.65	1.65	1.64	1.64

Table 2.1: Comparison of β quantiles of $|\widetilde{I}_n|$ for different sample sizes n and β quantiles of standard half-normal distribution

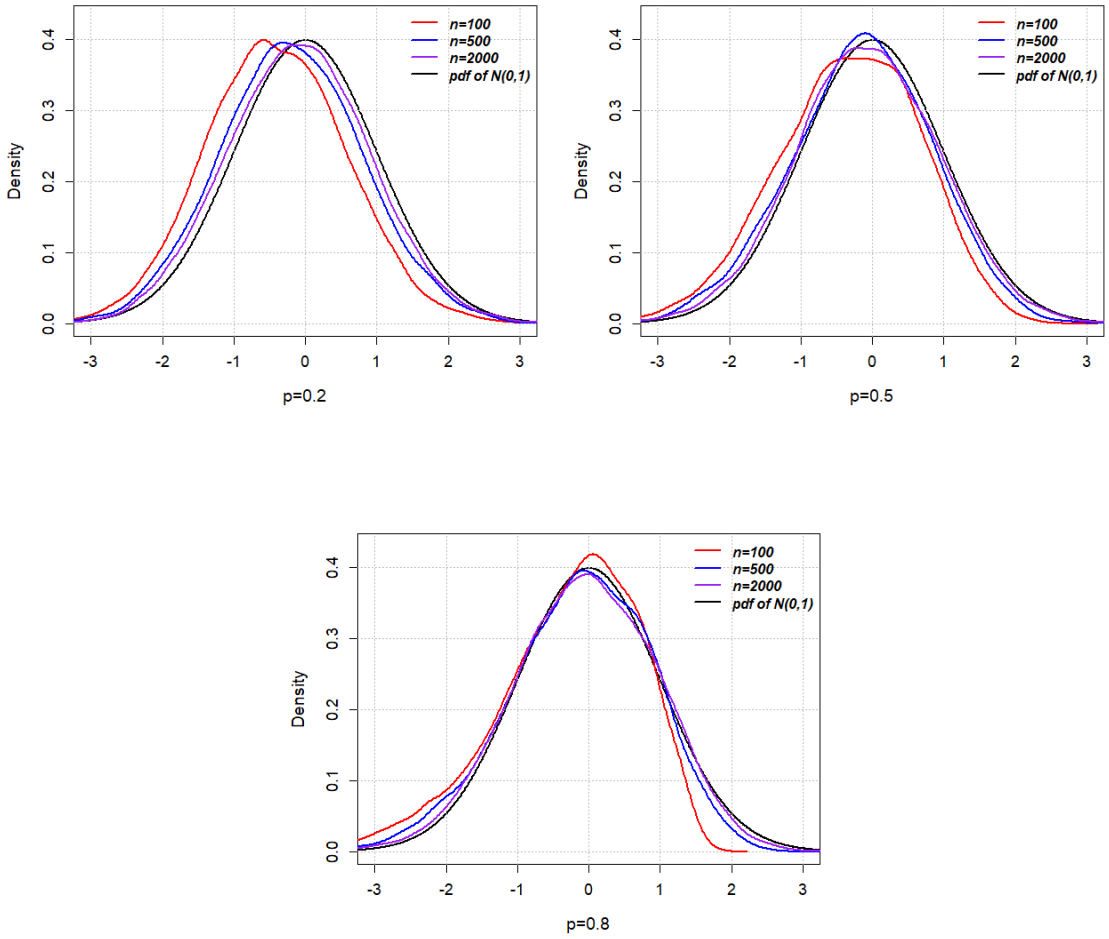


Figure 2.1: Density estimates of the statistic $\widetilde{I}_n$ across various sample sizes n and parameter values p

by means of a parametric bootstrap procedure based on the maximum likelihood estimator of p, given by $\hat{p}_n = \bar{X}_n^{-1}$. In addition to the newly proposed tests, the power study includes several well-established competitors, as well as some more recently introduced ones, which are presented below:

- the test statistic proposed in Klar, 1999, given by

$$K_n = \sup_{t \geq 0} |\Psi_n(t) - \hat{\Psi}(t)|,$$

where $\Psi(t) = \int_t^\infty (1 - F(x;p))dx$, $\Psi_n(t)$ is its empirical counterpart and $\hat{\Psi}(t)$ is obtained by replacing $F(x;p)$ with $F(x;\hat{p}_n)$ in the expression for $\Psi(t)$. Here, $F(x;p)$ denotes the cumulative distribution function of the geometric law (2.1);

- the test statistic proposed in Jiménez-Gamero et al., 2009, given by

$$W_n = \int_{\mathbb{R}} (\varphi_n(t) - \varphi(t;\hat{p}_n))^2 f(t)dt,$$

where $\varphi_n(t) = \frac{1}{n}\sum_{j=1}^{n} e^{itX_j}$ and $\varphi(t;p) = \frac{pe^{it}}{1-(1-p)e^{it}}$, $t \in \mathbb{R}$ denote, respectively, the empirical characteristic function and the theoretical characteristic function corresponding to the geometric law given in (2.1). Here, the weight function is chosen as $f(t) = \frac{1}{\pi(1+t^2)}$, which is one of the suitable choices suggested by the authors;

- the modified Cramér-von Mises test statistic, proposed in Henze, 1996, given by

$$H_n = \sum_{k=1}^{X_{(n)}} \{F_n(k) - F(k;\hat{p}_n)\}^2 \{F_n(k) - F_n(k-1)\},$$

where $F_n(x)$ is the empirical distribution, $F(x;p)$ is the cumulative distribution function of the geometric law (2.1) and $X_{(n)} = \max_{1 \leq i \leq n} X_i$;

- the test statistic based on a characterization of the geometric distribution through linear regression of order statistics, proposed in Jiménez-Gamero and Alba-Fernández, 2021, given by

$$L_n = \frac{1}{n^2(n-1)^2} \sum_{\substack{i,j,k,l=1 \\ i \neq j, k \neq l}}^{n} (|X_i - X_j| - \hat{a})(|X_k - X_l| - \hat{a}) e^{-\frac{(\min(X_i,X_j) - \min(X_k,X_l))^2}{2}},$$

where $\hat{a} = \frac{2\hat{q}_n}{1-\hat{q}_n^2}$ and $\hat{q}_n = 1 - \hat{p}_n$;

- the test statistics, proposed in Milošević et al., 2021, which are based on a

characterization of the geometric distribution through a differential equation satisfied by its pgf given by

$$T_n^{(2)} = \int_0^1 S_n^2(t;\hat{p}_n)w(t)dt \text{ and } T_n^{(3)} = \sup_{t\in[0,1]} |S_n(t;\hat{p}_n)|,$$

where $S_n(t;p) = pt^2 g_n'(t) - g_n^2(t)$, g_n is the empirical pgf and $w(t) = t$ is one of the weight functions considered by the authors and is adopted here, since it exhibited slightly better performance in their reported results.

For all competitors, the null hypothesis is rejected for large values of the associated test statistic.

The accuracy of the bootstrap approximation to the null distributions of the test statistics under consideration is assessed through an analysis of the corresponding empirical sizes. The study is conducted at the significance level $\alpha = 0.05$ for the sample sizes $n = 20$ and $n = 40$, based on samples generated from the geometric distribution $\mathcal{G}(p)$ with $p = 0.1, 0.2, \ldots, 0.9$. For each configuration, the test statistics are computed and the associated bootstrap p-values are obtained using $B = 1000$ bootstrap replications, with the entire experiment repeated $N = 1000$ times.

The results presented in Table 2.2 suggest that the quality of the parametric bootstrap approximation is strongly influenced by the value of p. For small and moderate values of p, the empirical rejection probabilities remain close to the nominal significance level. In contrast, for larger values of p, the resulting tests tend to be conservative. The only exception is observed for $n = 20$ and $p = 0.9$, where the tests become somewhat liberal. In light of these finite-sample effects, particularly for smaller sample sizes, the use of mid-p values may represent a possible improvement; see Wilson and Einbeck, 2023.

To assess power performance, several commonly used alternative scenarios are examined. More precisely, the analysis is carried out under the following alternative distributions:

- a shifted Poisson distribution P(λ)

$$\mathbb{P}\{X = k\} = \frac{\lambda^{k-1}e^{-\lambda}}{(k-1)!}, \;\; k \in \mathbb{N}, \; \lambda > 0;$$

	$n=20$								$n=40$							
p	T_n	I_n	K_n	W_n	H_n	$T_n^{(2)}$	$T_n^{(3)}$	L_n	T_n	I_n	K_n	W_n	H_n	$T_n^{(2)}$	$T_n^{(3)}$	L_n
0.1	5	6	4	6	5	6	6	6	6	5	6	6	6	7	7	5
0.2	5	5	3	4	4	5	5	4	5	5	4	5	5	4	5	4
0.3	6	6	4	5	5	6	6	5	5	5	4	5	5	5	5	5
0.4	6	6	3	6	5	5	5	5	6	6	4	5	5	5	5	5
0.5	6	6	3	5	5	4	4	3	5	5	5	6	6	6	5	4
0.6	5	5	3	5	6	4	4	3	4	4	3	5	5	4	4	3
0.7	4	5	3	5	5	4	3	2	5	5	4	4	5	3	3	3
0.8	4	5	4	5	5	4	5	2	4	4	4	5	5	3	3	1
0.9	12	12	15	15	15	15	15	12	2	2	6	5	6	5	5	2

Table 2.2: Empirical type I error (percentage of rejections of the null hypothesis) at significance level $\alpha = 0.05$ for sample sizes $n = 20$ and $n = 40$

- a shifted binomial distribution $\mathrm{B}(n, p)$

$$\mathbb{P}\{X = k\} = \binom{n}{k-1} p^{(k-1)}(1-p)^{n-k+1}, \;\; k = 1, ..., n+1, \; n \in \mathbb{N}, \; p \in (0,1);$$

- a shifted negative binomial distribution $\mathrm{NB}(n, p)$

$$\mathbb{P}\{X = k\} = \frac{(n+k-2)!}{(n-1)!(k-1)!} p^n (1-p)^{k-1}, \;\; k \in \mathbb{N}, \; n \in \mathbb{N}, \; p \in (0,1);$$

- a type 1 discrete Weibull distribution $\mathrm{W}_1(b, q)$

$$\mathbb{P}\{X = k\} = q^{(k-1)^b} - q^{k^b}, \;\; k \in \mathbb{N}, \; q \in (0,1), \; b > 0;$$

- a shifted type 3 discrete Weibull distribution $\mathrm{W}_3(q, b)$

$$\mathbb{P}\{X = k\} = e^{-q \sum_{i=1}^{k-1} i^b} \left(1 - e^{-qk^b}\right), \;\; k \in \mathbb{N}, \; q > 0, \; b \geq -1,$$

letting $\sum_{i=1}^{k-1} i^b = 0$ if $k = 1$;

- a law of geometric - discrete uniform mixture distribution $\mathrm{MixGU}(\omega, p)$,

$$\omega X + (1-\omega)Y, \;\; 0 < \omega < 1,$$

where X and Y are independent, $X \sim G(p)$ and $Y \sim DU(1, 2)$;

- a law of shifted absolute difference between geometric and discrete uniform distribution $\mathrm{AbsGU}(p)$,

$$|X - Y| + 1,$$

		$n = 20$								$n = 40$							
Alt.	p_0	T_n	I_n	K_n	W_n	H_n	$T_n^{(2)}$	$T_n^{(3)}$	L_n	T_n	I_n	K_n	W_n	H_n	$T_n^{(2)}$	$T_n^{(3)}$	L_n
P(0.5)	0.67	22	**23**	12	15	16	6	5	10	40	**42**	26	30	30	23	22	27
P(1)	0.5	**46**	**46**	27	37	31	35	31	38	82	**83**	66	70	68	72	70	74
P(1.5)	0.4	53	53	49	60	42	62	59	**63**	93	**94**	86	87	85	92	92	93
P(2)	0.33	50	50	71	75	59	**81**	79	**81**	94	94	98	97	97	**99**	98	**99**
B(5,0.2)	0.5	**73**	**73**	53	61	54	59	55	61	95	**96**	90	90	89	92	91	92
B(5,0.05)	0.8	10	**11**	3	7	7	1	1	2	16	**17**	9	11	12	4	3	5
B(10,0.05)	0.67	28	**29**	15	20	21	10	8	15	**44**	**44**	31	33	33	25	23	31
B(10,0.1)	0.5	**57**	**57**	39	48	42	46	43	48	88	**89**	76	78	77	80	79	83
NB(2,0.2)	0.11	11	10	21	22	17	**33**	32	32	28	27	48	44	47	**65**	63	**65**
NB(2,0.4)	0.25	17	17	12	19	9	20	19	**21**	36	37	34	36	29	43	41	**44**
NB(2,0.6)	0.43	**17**	**17**	9	13	8	12	11	13	**30**	**30**	16	21	16	20	18	22
NB(2,0.8)	0.67	**11**	**11**	5	8	9	3	3	4	12	**13**	7	9	9	5	5	6
W_1(0.5,0.8)	0.02	60	50	82	74	**95**	**95**	93	92	96	95	99	96	**100**	**100**	**100**	**100**
W_1(1.2,0.8)	0.26	**14**	13	7	11	5	12	11	12	21	**22**	15	16	13	21	20	21
W_1(1.4,0.6)	0.5	**33**	**33**	18	26	19	21	19	25	**59**	**59**	39	45	42	45	43	49
W_1(1.4,0.8)	0.32	26	27	19	27	14	30	28	**31**	55	56	49	48	43	60	58	**62**
W_3(0.2,0.5)	0.28	23	23	18	22	12	**25**	24	**25**	51	**52**	43	41	35	**52**	51	**52**
W_3(0.1,0.75)	0.22	34	33	43	44	32	**58**	56	**58**	70	71	84	76	78	**90**	**90**	89
W_3(0.4,0.5)	0.42	**23**	**23**	13	21	13	20	18	21	49	**50**	33	34	29	38	36	42
W_3(0.8,0.5)	0.61	21	**22**	11	16	14	10	8	13	34	**35**	20	24	23	19	18	23
MixGU(0.5,0.5)	0.57	**35**	33	18	30	31	16	14	19	**48**	45	28	46	42	20	19	28
MixGU(0.5,0.7)	0.68	42	**43**	25	35	37	15	13	20	**69**	68	53	61	60	41	39	49
MixGU(0.3,0.6)	0.64	**64**	63	43	54	56	29	26	39	**86**	84	71	80	78	56	55	68
MixGU(0.6,0.5)	0.55	**20**	19	10	15	16	8	8	10	**32**	30	19	29	27	14	14	19
AbsGU(0.3)	0.32	4	4	4	**17**	9	7	5	7	5	4	5	**28**	17	7	7	8
AbsGU(0.5)	0.5	40	38	23	**44**	39	22	21	30	69	67	44	**70**	66	37	34	53
AbsGU(0.7)	0.61	**76**	**76**	59	70	70	47	44	58	**97**	**97**	91	95	94	84	82	90

Table 2.3: Percentage of rejected null hypotheses at the significance level of 0.05 for sample sizes $n = 20$ and $n = 40$

where X and Y are independent, $X \sim G(p)$ and $Y \sim DU(1,2)$.

Table 2.3 presents the results, with the highest powers highlighted in bold. To facilitate interpretation, the table also includes the quantity $p_0 = \frac{1}{E(X)}$, which represents the value of the geometric parameter obtained by matching the mean of the alternative distribution with that of a geometric distribution. The results indicate that no single test is uniformly more powerful over all considered alternatives, which agrees with the findings of Janssen, 2000. Taken as a whole, the proposed tests exhibit competitive performance compared to the remaining tests. The newly proposed tests generally outperform competing tests for moderate and large values of p_0. However, for alternatives that correspond to small values of p_0, their performance is less satisfactory. In particular, the proposed tests exhibit similar power performance across all cases examined, as expected, since they leverage the same characterizing property. What may be regarded as surprising is their distinct behavior in comparison with L_n, although L_n is based on a characterization related to Ferguson's. This difference could be related to the fact that L_n is based on the regression-type condition $E(|X_2 - X_1| \mid \min(X_1, X_2) = j) = a,\ \forall j \geq 1$ where a denotes the corresponding value of $E(|X_2 - X_1|)$ under the geometric model. By contrast, the statistic considered here relies on the full independence between $|X_1 - X_2|$ and $\min(X_1, X_2)$.

Chapter 3

A correlation-type goodness-of-fit test for absolutely continuous distributions

Consider i.i.d random variables $X_1, X_2, \ldots, X_n$ defined on the probability space $(\Omega, \mathcal{A}, P)$. Their common, unknown distribution is denoted by F, which is assumed to belong to a broad nonparametric class of probability measures

$$\mathcal{G} = \{G : G \text{ is a probability measure on } (\mathbb{R}, \mathcal{B}(\mathbb{R}))\}.$$

Let $\mathcal{P} = \{F_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \Theta\}$ be a parametric family indexed by a finite-dimensional parameter taking values in $\Theta \subset \mathbb{R}^d$, for some fixed integer d. Assume that $\mathcal{P}$ is a proper subfamily of $\mathcal{G}$.

The objective is to determine whether the true underlying distribution F belongs to the specified parametric family. Formally, the following hypothesis testing problem is considered:

$$H_0 : F \in \mathcal{P}, \tag{3.1}$$

against the alternative hypothesis

$$H_1 : F \in \mathcal{G} \setminus \mathcal{P}.$$

As already noted, among the earliest and most established approaches to testing null hypotheses of this type are the classical edf tests, such as

- Kolmogorov-Smirnov test

$$\mathrm{KS}_n = \sqrt{n} \sup_{x \in \mathbb{R}} |F_n(x) - F_{\hat{\boldsymbol{\theta}}_n}(x)|, \tag{3.2}$$

· Cramér-von Mises test

$$\mathrm{CvM}_n = \int_{-\infty}^{\infty} (F_n(x) - F_{\hat{\boldsymbol{\theta}}_n}(x))^2 dF_{\hat{\boldsymbol{\theta}}_n}(x), \tag{3.3}$$

· Anderson-Darling test

$$\mathrm{AD}_n = \int_{-\infty}^{\infty} \frac{(F_n(x) - F_{\hat{\boldsymbol{\theta}}_n}(x))^2}{F_{\hat{\boldsymbol{\theta}}_n}(x)(1 - F_{\hat{\boldsymbol{\theta}}_n}(x))} dF_{\hat{\boldsymbol{\theta}}_n}(x).$$

Here, F_n represents the empirical distribution function of the sample $X_1, \ldots, X_n$ and $\hat{\boldsymbol{\theta}}_n = \hat{\boldsymbol{\theta}}_n(X_1, \ldots, X_n)$ is an estimator of the unknown parameter $\boldsymbol{\theta}$ under the null hypothesis. Indeed, these tests rely on the difference between the empirical distribution function of the data and a parametric estimator of the theoretical distribution function under the null hypothesis (see, e.g., D'Agostino and Stephens, 1986 and Stute et al., 1993).

Another approach relies on distributional characterizations, where the null model is identified through a property that is unique to the corresponding family. In this context, attention is focused on a particular class of distributional characterizations, namely independence characterizations. In particular, several gof tests based on this type of characterization have been proposed for specific parametric families (see, e.g., Mudholkar et al., 2001; Milošević, 2017; Jiménez-Gamero et al., 2020).

In the general case, Milošević and Obradović, 2016 developed a method for constructing Kolmogorov-type test statistics derived from this class of characterizations. In contrast, Baringhaus and Gaigall, 2015 proposed an integral-type test statistic arising from the same characterization-based approach. This test statistic is based on an L^2-type distance between the joint U-empirical distribution function of the two random variables appearing in the characterization, say $\psi_1(X, Y)$ and $\psi_2(X, Y)$, and the product of its marginals. Specifically, let $X_1, \ldots, X_n$ be an i.i.d. sample from the distribution of a random variable X, and let Y denote an independent copy of X. They consider the following test statistic,

$$\mathrm{HBKR}_n = n \int \Big(H_n(x, y) - F_n(x) G_n(y)\Big)^2 dH_n(x, y), \tag{3.4}$$

where

$$H_n(x,y) = \frac{1}{n(n-1)} \sum_{\substack{i,j=1 \\ i \neq j}}^{n} \mathbf{1}\{\psi_1(X_i, X_j) \leq x, \psi_2(X_i, X_j) \leq y\}$$

denotes the joint U-empirical distribution function of $(\psi_1(X,Y), \psi_2(X,Y))$, while $F_n(x)$ and $G_n(y)$ represent its marginals. The test statistic HBKR_n is a Hoeffding-Blum-Kiefer-Rosenblatt-type statistic adapted to the problem of testing the independence of $\psi_1(X,Y)$ and $\psi_2(X,Y)$.

Having reviewed some of the existing approaches to testing (3.1), attention is now turned to the development of a novel general testing procedure. The proposed test statistic is

$$T_n = \frac{\int H_n(x,y) F_n(x) G_n(y) dH_n(x,y)}{\sqrt{\int H_n^2(x,y) dH_n(x,y)} \sqrt{\int F_n^2(x) G_n^2(y) dH_n(x,y)}} - 1. \tag{3.5}$$

The motivation for this test statistic is as follows. First, let denote the joint distribution function of $(\psi_1(X,Y), \psi_2(X,Y))$ with

$$H(x,y) = \mathbb{P}\{\psi_1(X,Y) \leq x, \psi_2(X,Y) \leq y\},$$

where the marginal functions are given by:

$$\begin{aligned} F(x) &= \mathbb{P}\{\psi_1(X,Y) \leq x\} \text{ and} \\ G(y) &= \mathbb{P}\{\psi_2(X,Y) \leq y\}. \end{aligned}$$

Using the Cauchy-Schwarz inequality and the nonnegativity of the distribution function, the following is obtained

$$\begin{aligned} 0 &\leq \int H(x,y) F(x) G(y) dH(x,y) \\ &\leq \sqrt{\int H^2(x,y) dH(x,y)} \sqrt{\int F^2(x) G^2(y) dH(x,y)}. \end{aligned}$$

The equality is reached iff $H(x,y) = F(x)G(y),\ \forall x, y \in \mathbb{R}$ such that the joint distribution function $H(x,y)$ is nonzero, which is a condition that is fulfilled solely under the null hypothesis. Therefore, it holds

$$\frac{\int H(x,y) F(x) G(y) dH(x,y)}{\sqrt{\int H^2(x,y) dH(x,y)} \sqrt{\int F^2(x) G^2(y) dH(x,y)}} \leq 1.$$

The proposed test statistic is obtained by replacing the theoretical distribution function with the empirical one in the expression above.

Since the test statistics HBKR_n and T_n are closely related, one might expect that tests based on these statistics would exhibit similar power. In the following, a comparative analysis under various alternatives is conducted to examine whether the tests exhibit similar power.

Throughout this chapter, only two-dimensional sample functions ψ_1 and ψ_2 are considered. However, the results admit a natural extension to the case where the family $\mathcal{P}$ is characterized by the independence of $\psi_1(X_1, \ldots, X_d)$ and $\psi_2(X_1, \ldots, X_d)$ for some $d \geq 2$.

3.1 Large sample properties

Let H denote the absolutely continuous joint distribution function of $(\psi_1(X, Y), \psi_2(X, Y))$, and let H_n be its empirical counterpart. By Theorem 1.13 and the discussion following it, with $m = r = 2$, it holds

$$\sup_{(x,y)\in\overline{\mathbb{R}}^2} |H_n(x,y) - H(x,y)| \xrightarrow[n\to\infty]{a.s.} 0. \tag{3.6}$$

Prior to establishing the almost sure limit of the proposed test statistic, the auxiliary lemma is introduced.

Lemma 3.1 (Halaj et al., 2024)**.** *Assume that H is an absolutely continuous two-dimensional df, then:*

(a) $\int H(x,y)F(x)G(y)dH_n(x,y) \xrightarrow[n\to\infty]{a.s.} \int H(x,y)F(x)G(y)dH(x,y)$

(b) $\int H^2(x,y)dH_n(x,y) \xrightarrow[n\to\infty]{a.s.} \int H^2(x,y)dH(x,y)$

(c) $\int F^2(x)G^2(y)dH_n(x,y) \xrightarrow[n\to\infty]{a.s.} \int F^2(x)G^2(y)dH(x,y)$

(d) $\int H_n^2(x,y)dH_n(x,y) \xrightarrow[n\to\infty]{a.s.} \int H^2(x,y)dH(x,y)$

(e) $\int F_n^2(x)G_n^2(y)dH_n(x,y) \xrightarrow[n\to\infty]{a.s.} \int F^2(x)G^2(y)dH(x,y)$

Proof. The proof of (a)-(c) is a direct consequence of (3.6) and the Portmanteau Lemma (see Lemma 1.1). In case (d) one has

$$\int H_n^2(x,y)dH_n(x,y) = \int H^2(x,y)dH_n(x,y) + \int (H_n^2(x,y) - H^2(x,y))dH_n(x,y). \tag{3.7}$$

From part (b), it follows

$$\int H^2(x,y)dH_n(x,y) \xrightarrow[n\to\infty]{a.s.} \int H^2(x,y)dH(x,y). \tag{3.8}$$

Since

$$\begin{aligned}|H_n^2(x,y) - H^2(x,y)| =& |H_n(x,y) - H(x,y)|\,|H_n(x,y) + H(x,y)| \\ \leq& 2|H_n(x,y) - H(x,y)|,\end{aligned}$$

from (3.6) it readily follows that

$$\int (H_n^2(x,y) - H^2(x,y))dH_n(x,y) \xrightarrow[n\to\infty]{a.s.} 0. \tag{3.9}$$

From (3.8) and (3.9), it directly follows that part (d) holds. In part (e), by following the same approach

$$\begin{aligned}\int F_n^2(x)G_n^2(y)dH_n(x,y) =& \int F^2(x)G^2(y)dH_n(x,y) \\ &+ \int (F_n^2(x) - F^2(x))G^2(y)dH_n(x,y) \\ &+ \int (F_n^2(x) - F^2(x))(G_n^2(y) - G^2(y))dH_n(x,y) \\ &+ \int (G_n^2(y) - G^2(y))F^2(x)dH_n(x,y)\end{aligned}$$

all terms except the first one converge *a.s.* to zero. The result follows using part (c) for the first term. □

Remark 3.1. *Under the null hypothesis, all limits in Lemma 3.1 are equal to* $\left(\int F^2(x)dF(x)\right)^2 = 1/9$*, so it follows that*

$$\sqrt{\int H_n^2(x,y)dH_n(x,y)}\sqrt{\int F_n^2(x)G_n^2(y)dH_n(x,y)} \xrightarrow[n\to\infty]{a.s.} 1/9.$$

Now, the *a.s.* limit of the proposed test statistic under general distributional assumptions is considered.

Theorem 3.1 (Halaj et al., 2024)**.** *Suppose that* H*, the joint distribution function of* $(\psi_1(X,Y),\ \psi_2(X,Y))$*, is absolutely continuous, then*

$$T_n \xrightarrow[n\to\infty]{a.s.} T = \frac{\int H(x,y)F(x)G(y)dH(x,y)}{\sqrt{\int H^2(x,y)dH(x,y)}\sqrt{\int F^2(x)G^2(y)dH(x,y)}} - 1.$$

Proof. The result follows directly from Lemma 3.1. □

In the preceding section, it was established that $T \leq 0$ with equality holding under the null hypothesis. Moreover, $T = 0$ iff H_0 is true. Therefore, a reasonable test should reject the null hypothesis for small values of the test statistic T_n. In order to determine the corresponding critical values, the distribution of T_n under the null hypothesis is required. Since the null distribution of T_n is unknown, a suitable approximation is needed. For this purpose, its asymptotic null distribution is derived.

First, the necessary notation is introduced. Let $\overline{\mathbb{R}}^2 = [-\infty, \infty]^2$, which is a compact space when it is endowed with the metric

$$\rho((x_1, y_1), (x_2, y_2)) = \max\{|\arctan(x_1) - \arctan(x_2)|, |\arctan(y_1) - \arctan(y_2)|\},$$

for all $(x_1, y_1), (x_2, y_2) \in \overline{\mathbb{R}}^2$. Let $\ell^\infty(\overline{\mathbb{R}}^2)$ denotes the space of all bounded real-valued functions on $\overline{\mathbb{R}}^2$, equipped with the supremum norm

$$\|f\|_\infty = \sup_{(x,y)\in\overline{\mathbb{R}}^2} |f(x, y)|.$$

Consider the process $\mathbb{W}_n = \{W_n(x, y),\ (x, y) \in \overline{\mathbb{R}}^2\}$, where

$$W_n(x, y) = \sqrt{n}\,\{H_n(x, y) - H(x, y)\},\ \ (x, y) \in \overline{\mathbb{R}}^2.$$

By Theorem 1 of Baringhaus and Gaigall, 2015 and the subsequent discussion, there exists a zero-mean Gaussian process $\mathbb{W} = \{W(x, y),\ (x, y) \in \overline{\mathbb{R}}^2\}$ that has $\mathbb{P}$-*a.s.* uniformly ρ-continuous sample paths, and that can be regarded as a tight Borel measurable random element in $\ell^\infty(\overline{\mathbb{R}}^2)$ such that

$$\mathbb{W}_n \xrightarrow[n\to\infty]{\mathcal{D}} \mathbb{W}. \tag{3.10}$$

The statement that the process $\mathbb{W}$ has $\mathbb{P}$-*a.s.* uniformly ρ-continuous sample paths means that there exists a set Ω_0 with $\mathbb{P}(\Omega_0) = 1$ such that, for every $\omega \in \Omega_0$, the function

$$(x, y) \mapsto \mathbb{W}(x, y; \omega)$$

is bounded and uniformly continuous with respect to the metric ρ. More precisely, for every $\varepsilon > 0$, there exists $\delta > 0$ such that, for all $(x_1, y_1), (x_2, y_2) \in \overline{\mathbb{R}}^2$,

$$\rho\big((x_1, y_1), (x_2, y_2)\big) < \delta \quad \Longrightarrow \quad |\mathbb{W}(x_1, y_1; \omega) - \mathbb{W}(x_2, y_2; \omega)| < \varepsilon.$$

In addition, let $\mathbb{U}_n = \{U_n(x,y), (x,y) \in \overline{\mathbb{R}}^2\}$, where

$$U_n(x,y) = \sqrt{n}\left\{H_n(x,y) - F_n(x)G_n(y)\right\}, \quad (x,y) \in \overline{\mathbb{R}}^2.$$

The corresponding weak convergence result for the process $\mathbb{U}_n$ is stated in the following theorem, provided in Baringhaus and Gaigall, 2015.

Theorem 3.2 (Baringhaus and Gaigall, 2015)**.** *Assume that the null hypothesis is true and that the marginal distribution functions of $(\psi_1(X_1,X_2), \psi_2(X_1,X_2))$ are continuous. Then, there exists a zero-mean Gaussian process $\mathbb{U} = \{U(x,y) \mid (x,y) \in \overline{\mathbb{R}}^2\}$ that has $\mathbb{P}$-a.s. uniformly ρ-continuous sample paths, and that can be regarded as a tight Borel measurable random element in $\ell^\infty(\overline{\mathbb{R}}^2)$ such that $\mathbb{U}_n \xrightarrow[n\to\infty]{\mathcal{D}} \mathbb{U}$. The covariance function C of $\mathbb{U}$ is given by*

$$\begin{aligned}
&C((x,y),(\tilde{x},\tilde{y})) \\
&\quad= \sum_{(i,j)\in J} \mathbb{P}\{\psi_1(X_1,X_2) \le x, \psi_2(X_1,X_2) \le y, \psi_1(X_i,X_j) \le \tilde{x}, \psi_2(X_i,X_j) \le \tilde{y}) \\
&\quad -G(\tilde{y}) \sum_{(i,j)\in J} \mathbb{P}\{\psi_1(X_1,X_2) \le x, \psi_2(X_1,X_2) \le y, \psi_1(X_i,X_j) \le \tilde{x}) \\
&\quad -F(\tilde{x}) \sum_{(i,j)\in J} \mathbb{P}\{\psi_1(X_1,X_2) \le x, \psi_2(X_1,X_2) \le y, \psi_2(X_i,X_j) \le \tilde{y}) \\
&\quad -G(y) \sum_{(i,j)\in J} \mathbb{P}\{\psi_1(X_1,X_2) \le x, \psi_1(X_i,X_j) \le \tilde{x}, \psi_2(X_i,X_j) \le \tilde{y}) \\
&\quad -F(x) \sum_{(i,j)\in J} \mathbb{P}\{\psi_2(X_1,X_2) \le y, \psi_1(X_i,X_j) \le \tilde{x}, \psi_2(X_i,X_j) \le \tilde{y}) \\
&\quad +G(y)G(\tilde{y}) \sum_{(i,j)\in J} \mathbb{P}\{\psi_1(X_1,X_2) \le x, \psi_1(X_i,X_j) \le \tilde{x}) \\
&\quad +G(y)F(\tilde{x}) \sum_{(i,j)\in J} \mathbb{P}\{\psi_1(X_1,X_2) \le x, \psi_2(X_i,X_j) \le \tilde{y}) \\
&\quad +F(x)G(\tilde{y}) \sum_{(i,j)\in J} \mathbb{P}\{\psi_1(X_i,X_j) \le \tilde{x}, \psi_2(X_1,X_2) \le y) \\
&\quad +F(x)F(\tilde{x}) \sum_{(i,j)\in J} \mathbb{P}\{\psi_2(X_1,X_2) \le y, \psi_2(X_i,X_j) \le \tilde{y}) \\
&\quad -4F(x)G(y)F(\tilde{x})G(\tilde{y}), \quad (x,y),(\tilde{x},\tilde{y}) \in \overline{\mathbb{R}}^2,
\end{aligned}$$

where $J = \{(1,3),(3,1),(2,3),(3,2)\}$.

To facilitate the proof of the asymptotic distribution of the proposed test statistic under the null hypothesis, two auxiliary lemmas are presented first.

Lemma 3.2 (Halaj et al., 2024)**.** *Suppose that the null hypothesis is true and that the marginal distribution functions F and G are continuous. Let* $\mathbb{V}_n = \{V_n(x,y),\ (x,y) \in \overline{\mathbb{R}}^2\}$*, where*

$$V_n(x,y) = \sqrt{n}\left\{F_n(x)G_n(y) - F(x)G(y)\right\}.$$

There is a zero-mean Gaussian process $\mathbb{V} = \{V(x,y),\ (x,y) \in \overline{\mathbb{R}}^2\}$ *that has P-a.s. uniformly* ρ*-continuous sample paths, and that can be regarded as a tight Borel measurable random element in* $\ell^\infty(\overline{\mathbb{R}}^2)$ *such that* $\mathbb{V}_n \xrightarrow[n\to\infty]{\mathcal{D}} \mathbb{V}$*.*

Proof. The result can be derived using the same arguments as in the proof of Theorem 3.2 (presented in Baringhaus and Gaigall, 2015). First, rewrite the process $\mathbb{V}_n$ as $\mathbb{V}_n = \phi(\mathbb{W}_n) + \frac{1}{\sqrt{n}}\chi(\mathbb{W}_n)$, where the functions $\phi, \chi : \ell^\infty(\overline{\mathbb{R}}^2) \to \ell^\infty(\overline{\mathbb{R}}^2)$ defined by

$$\begin{aligned}\phi(w)(x,y) &= G(y)w(x,\infty) + F(x)w(\infty,y),\\ \chi(w)(x,y) &= w(x,\infty)w(\infty,y),\end{aligned}$$

for all $(x,y) \in \overline{\mathbb{R}}^2$ and $w \in \ell^\infty(\overline{\mathbb{R}}^2)$, are continuous and measurable with respect to the σ-algebra generated by the projections on $\ell^\infty(\overline{\mathbb{R}}^2)$. By the continuous mapping theorem (Theorem 1.2), together with

$$\frac{1}{\sqrt{n}}\chi(\mathbb{W}_n) \xrightarrow[n\to\infty]{\mathbb{P}} 0,$$

it follows that

$$\mathbb{V}_n \xrightarrow[n\to\infty]{\mathcal{D}} \phi(\mathbb{W}).$$

Denote the limiting process by $\mathbb{V} = \phi(\mathbb{W})$. Since ϕ is a continuous linear mapping, $\mathbb{V}$ has bounded and $\mathbb{P}$-*a.s.* uniformly ρ-continuous sample paths. Furthermore, as a linear transformation of the zero-mean Gaussian process $\mathbb{W}$, the process $\mathbb{V}$ is also a zero-mean Gaussian. □

Lemma 3.3 (Halaj et al., 2024)**.** *Suppose that the null hypothesis is true and that the marginal distribution functions F and G are continuous. Then*

(a) $\sqrt{n}\int H(x,y)\{H_n(x,y) - H(x,y)\}dH_n(x,y) \xrightarrow[n\to\infty]{\mathcal{D}} \int H(x,y)W(x,y)dH(x,y).$

(b) $\sqrt{n}\int H(x,y)\{H_n(x,y) - F_n(x)G_n(y)\}dH_n(x,y) \xrightarrow[n\to\infty]{\mathcal{D}} \int H(x,y)U(x,y)dH(x,y).$

(c) $\sqrt{n}\int H(x,y)\{F_n(x)G_n(y) - H(x,y)\}dH_n(x,y) \xrightarrow[n\to\infty]{\mathcal{D}} \int H(x,y)V(x,y)dH(x,y).$

(d) $n\int \{H_n(x,y) - H(x,y)\}^2 dH_n(x,y) \xrightarrow[n\to\infty]{\mathcal{D}} \int W^2(x,y)dH(x,y).$

(e) $n\int\{F_n(x)G_n(y) - H(x,y)\}^2 dH_n(x,y) \xrightarrow[n\to\infty]{\mathcal{D}} \int V^2(x,y)dH(x,y).$

Proof. The main ideas of the proof are related to those used in the proof of Theorem 3 in Baringhaus and Gaigall, 2015. However, the arguments are adapted to the present framework and to the specific form of the proposed test statistic. For this reason, only part (a) is treated in detail; the proofs of the remaining parts proceed along the same lines.

Let $\mathcal{P}$ denote the space of probability measures on $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$ under the topology of weak convergence. The corresponding Borel σ-algebra is denoted by $\mathcal{B}(\mathcal{P})$. The empirical distribution H_n is regarded as a $\mathcal{B}(\mathcal{P})$-measurable random element of $\mathcal{P}$. By Varadarajan's theorem (Theorem 1.1),

$$H_n \xrightarrow[n\to\infty]{a.s.} H \quad \text{in } \mathcal{P},$$

where H is understood as a non-random element of $\mathcal{P}$. Hence,

$$H_n \xrightarrow[n\to\infty]{\mathcal{D}} H \quad \text{in } \mathcal{P}.$$

Furthermore, by (3.10),

$$\mathbb{W}_n \xrightarrow[n\to\infty]{\mathcal{D}} \mathbb{W} \quad \text{in } \ell^\infty(\overline{\mathbb{R}}^2).$$

Let u_ρ denote the space of bounded, uniformly ρ-continuous real-valued functions on $\overline{\mathbb{R}}^2$, equipped with the supremum norm, i.e.

$$u_\rho = \left\{\phi \in \ell^\infty(\overline{\mathbb{R}}^2) : \phi \text{ is uniformly } \rho\text{-continuous}\right\}.$$

Since $\mathbb{W}$ has $\mathbb{P}$-*a.s.* uniformly ρ-continuous sample paths, it follows that

$$\mathbb{P}(\mathbb{W} \in u_\rho) = 1.$$

Since $(\overline{\mathbb{R}}^2, \rho)$ is compact, u_ρ is separable under the supremum norm. Hence $\mathbb{W}$ is separable as an $\ell^\infty(\overline{\mathbb{R}}^2)$-valued random element. Since the limiting measure H is non-random and $\mathbb{W}$ is separable as an $\ell^\infty(\overline{\mathbb{R}}^2)$-valued random element, Slutsky's theorem (Theorem 1.3) yields

$$(\mathbb{W}_n, H_n) \xrightarrow[n\to\infty]{\mathcal{D}} (\mathbb{W}, H) \quad \text{in} \quad \ell^\infty(\overline{\mathbb{R}}^2) \times \mathcal{P}.$$

Consider now the mapping

$$\Phi(\phi,\mu) = \int H(x,y)\phi(x,y)\,d\mu(x,y), \quad (\phi,\mu) \in \ell^\infty(\overline{\mathbb{R}}^2) \times \mathcal{P}.$$

To justify the application of the continuous mapping theorem, it is enough to show that Φ is continuous at every point (ϕ,μ) such that $\phi \in u_\rho$. Let (ϕ_n,μ_n) be a sequence in $\ell^\infty(\overline{\mathbb{R}}^2) \times \mathcal{P}$ such that

$$(\phi_n,\mu_n) \to (\phi,\mu),$$

with respect to the product topology such that $\phi \in u_\rho$. In other words,

$$\|\phi_n - \phi\|_\infty \to 0$$

and

$$\mu_n \to \mu \ \text{ in } \ \mathcal{P}.$$

Then,

$$\left|\int H\phi_n\,d\mu_n - \int H\phi\,d\mu\right| \le \|H\|_\infty\|\phi_n-\phi\|_\infty + \left|\int H\phi\,d\mu_n - \int H\phi\,d\mu\right|.$$

The first term tends to zero by the uniform convergence of ϕ_n to ϕ. The second term tends to zero by the weak convergence of μ_n to μ in $\mathcal{P}$, since $H\phi$ is bounded and continuous. Thus, Φ is continuous at every point (ϕ,μ) with $\phi \in u_\rho$.

Since

$$\mathbb{P}(\mathbb{W} \in u_\rho) = 1,$$

the continuous mapping theorem gives

$$\Phi(\mathbb{W}_n, H_n) \xrightarrow[n\to\infty]{\mathcal{D}} \Phi(\mathbb{W}, H).$$

This gives the conclusion in part (a). □

Remark 3.2. *The limits obtained in Lemma 3.3 (a)-(c) are centered univariate normal random variables with finite variances. In contrast, the limits in parts (d) and (e) are linear combinations of independent chi-squared random variables with one degree of freedom, whose expectations are finite.*

The following result provides, in terms of the limiting process $\mathbb{U}$, the asymptotic null distribution of nT_n, where T_n is defined in (3.5).

Theorem 3.3 (Halaj et al., 2024)**.** *Suppose that the null hypothesis is true and that F and G are continuous. Then*

$$nT_n \xrightarrow[n\to\infty]{\mathcal{D}} \frac{81}{2}\left\{\int H(x,y)U(x,y)dH(x,y)\right\}^2 - \frac{9}{2}\int U^2(x,y)dH(x,y).$$

Proof. Using Remark 3.1 the denominator of T_n converges *a.s.* to $\frac{1}{9}$. Consider the numerator of T_n, denoted by N_n.

$$\begin{aligned} N_n = &\int \big(H_n(x,y)F_n(x)G_n(y) - H(x,y)F(x)G(y)\big)dH_n(x,y) \\ &+ \int H(x,y)F(x)G(y)dH_n(x,y) \\ &- \sqrt{\int H_n^2(x,y)dH_n(x,y)}\sqrt{\int F_n^2(x)G_n^2(y)dH_n(x,y)}. \end{aligned}$$

In order to shorten the notation, from now on, the arguments x and y of H, H_n, F, F_n, G_n and G will be suppressed. A direct calculation yields

$$\begin{aligned} N_n = &\int \big((H_n - H + H)(F_nG_n - H + H) - HFG\big)dH_n + \int HFG dH_n \\ &-\Big(\sqrt{\int H_n^2\,dH_n} - \sqrt{\int H^2\,dH_n} + \sqrt{\int H^2\,dH_n}\Big) \\ &\times\Big(\sqrt{\int F_n^2G_n^2\,dH_n} - \sqrt{\int H^2\,dH_n} + \sqrt{\int H^2\,dH_n}\Big) \\ &= \int H(H_n - H)dH_n + \int H(F_nG_n - H)dH_n + \int (H_n - H)(F_nG_n - H)dH_n \\ &-\Big(\sqrt{\int H_n^2\,dH_n} - \sqrt{\int H^2\,dH_n}\Big)\Big(\sqrt{\int F_n^2G_n^2\,dH_n} - \sqrt{\int H^2\,dH_n}\Big) \\ &-\sqrt{\int H^2\,dH_n}\Big(\sqrt{\int F_n^2G_n^2\,dH_n} - \sqrt{\int H^2\,dH_n}\Big) \\ &-\sqrt{\int H^2\,dH_n}\Big(\sqrt{\int H_n^2\,dH_n} - \sqrt{\int H^2\,dH_n}\Big) \end{aligned}$$

Taking into account relation $(a-b) = (\sqrt{a}-\sqrt{b})(\sqrt{a}+\sqrt{b})$, for $a, b \in \mathbb{R}$, $a, b \geq 0$, the following is obtained

$$\sqrt{\int H_n^2 dH_n} - \sqrt{\int H^2 dH_n} = \frac{\int (H_n^2 - H^2)dH_n}{\sqrt{\int H_n^2 dH_n} + \sqrt{\int H^2 dH_n}}. \tag{3.11}$$

Applying the identity $(a^2 - b^2) = (a-b)^2 + 2b(a-b)$, $a, b \in \mathbb{R}$ to the integrand

of the numerator in (3.11) leads to the expression

$$H_n^2 - H^2 = 2H(H_n - H) + (H_n - H)^2. \tag{3.12}$$

Now, it follows

$$\begin{aligned}
N_n &= \int H(H_n - H)dH_n + \int H(F_nG_n - H)dH_n + \int (H_n - H)(G_nF_n - H)dH_n \\
&\quad - \Big(\sqrt{\int H_n^2\, dH_n} - \sqrt{\int H^2\, dH_n}\Big)\Big(\sqrt{\int F_n^2G_n^2\, dH_n} - \sqrt{\int H^2\, dH_n}\Big) \\
&\quad - \sqrt{\int H^2\, dH_n}\frac{2\int H(F_nG_n - H)dH_n + \int (F_nG_n - H)^2 dH_n}{\sqrt{\int F_n^2G_n^2 dH_n} + \sqrt{\int H^2 dH_n}} \\
&\quad - \sqrt{\int H^2\, dH_n}\frac{2\int H(H_n - H)dH_n + \int (H_n - H)^2 dH_n}{\sqrt{\int H_n^2 dH_n} + \sqrt{\int H^2 dH_n}} \\
&= \int H(H_n - H)dH_n\frac{\sqrt{\int H_n^2 dH_n} - \sqrt{\int H^2 dH_n}}{\sqrt{\int H_n^2 dH_n} + \sqrt{\int H^2 dH_n}} \\
&\quad + \int H(F_nG_n - H)dH_n\frac{\sqrt{\int F_n^2G_n^2 dH_n} - \sqrt{\int H^2 dH_n}}{\sqrt{\int F_n^2G_n^2 dH_n} + \sqrt{\int H^2 dH_n}} \\
&\quad + \int (H_n - H)(G_nF_n - H)dH_n \\
&\quad - \Big(\sqrt{\int H_n^2\, dH_n} - \sqrt{\int H^2\, dH_n}\Big)\Big(\sqrt{\int F_n^2G_n^2\, dH_n} - \sqrt{\int H^2\, dH_n}\Big) \\
&\quad - \sqrt{\int H^2\, dH_n}\frac{\int (F_nG_n - H)^2 dH_n}{\sqrt{\int F_n^2G_n^2 dH_n} + \sqrt{\int H^2 dH_n}} \\
&\quad - \sqrt{\int H^2\, dH_n}\frac{\int (H_n - H)^2 dH_n}{\sqrt{\int H_n^2 dH_n} + \sqrt{\int H^2 dH_n}} \\
&= \int H(H_n - H)dH_n\frac{2\int H(H_n - H)dH_n + \int (H_n - H)^2 dH_n}{\Big(\sqrt{\int H_n^2 dH_n} + \sqrt{\int H^2 dH_n}\Big)^2} \\
&\quad + \int H(F_nG_n - H)dH_n\frac{2\int H(F_nG_n - H)dH_n + \int (F_nG_n - H)^2 dH_n}{\Big(\sqrt{\int F_n^2G_n^2 dH_n} + \sqrt{\int H^2 dH_n}\Big)^2} \\
&\quad + \int (H_n - H)(G_nF_n - H)dH_n \\
&\quad - \Big(\sqrt{\int H_n^2\, dH_n} - \sqrt{\int H^2\, dH_n}\Big)\Big(\sqrt{\int F_n^2G_n^2\, dH_n} - \sqrt{\int H^2\, dH_n}\Big)
\end{aligned}$$

$$
\begin{aligned}
&-\sqrt{\int H^2\,dH_n}\frac{\int (F_nG_n-H)^2dH_n}{\sqrt{\int F_n^2G_n^2dH_n}+\sqrt{\int H^2dH_n}}\\
&-\sqrt{\int H^2\,dH_n}\frac{\int (H_n-H)^2dH_n}{\sqrt{\int H_n^2dH_n}+\sqrt{\int H^2dH_n}}\\
=&2\left(\frac{\int H(H_n-H)dH_n}{\sqrt{\int H_n^2dH_n}+\sqrt{\int H^2dH_n}}-\frac{\int H(F_nG_n-H)dH_n}{\sqrt{\int F_n^2G_n^2dH_n}+\sqrt{\int H^2dH_n}}\right)^2\\
&-\sqrt{\int H^2\,dH_n}\frac{\int (F_nG_n-H)^2dH_n}{\sqrt{\int F_n^2G_n^2dH_n}+\sqrt{\int H^2dH_n}}\\
&-\sqrt{\int H^2\,dH_n}\frac{\int (H_n-H)^2dH_n}{\sqrt{\int H_n^2dH_n}+\sqrt{\int H^2dH_n}}\\
&+\int (H_n-H)(F_nG_n-H)dH_n+R_n,
\end{aligned}
$$

where

$$
\begin{aligned}
R_n=&\frac{\int H(H_n-H)dH_n\int (H_n-H)^2dH_n}{\left(\sqrt{\int H_n^2dH_n}+\sqrt{\int H^2dH_n}\right)^2}\\
&+\frac{\int H(F_nG_n-H)dH_n\int (F_nG_n-H)^2dH_n}{\left(\sqrt{\int F_n^2G_n^2dH_n}+\sqrt{\int H^2dH_n}\right)^2}\\
&-\frac{2\int H(H_n-H)dH_n\int (F_nG_n-H)^2dH_n}{\left(\sqrt{\int H_n^2dH_n}+\sqrt{\int H^2dH_n}\right)\left(\sqrt{\int F_n^2G_n^2dH_n}+\sqrt{\int H^2dH_n}\right)}\\
&-\frac{2\int (H_n-H)^2dH_n\int H(F_nG_n-H)dH_n}{\left(\sqrt{\int H_n^2dH_n}+\sqrt{\int H^2dH_n}\right)\left(\sqrt{\int F_n^2G_n^2dH_n}+\sqrt{\int H^2dH_n}\right)}\\
&-\frac{\int (H_n-H)^2dH_n\int (F_nG_n-H)^2dH_n}{\left(\sqrt{\int H_n^2dH_n}+\sqrt{\int H^2dH_n}\right)\left(\sqrt{\int F_n^2G_n^2dH_n}+\sqrt{\int H^2dH_n}\right)}.
\end{aligned}
$$

From Lemma 3.1 and Lemma 3.3, it follows that $nR_n=o_{\mathbb{P}}(1)$.
Now, from Lemma 3.1 (see also Remark 3.1),

$$
\begin{aligned}
\sqrt{\int H_n^2dH_n}+\sqrt{\int H^2dH_n}&\xrightarrow[n\to\infty]{a.s.}2/3,\\
\sqrt{\int F_n^2G_n^2dH_n}+\sqrt{\int H^2dH_n}&\xrightarrow[n\to\infty]{a.s.}2/3.
\end{aligned}
$$

Therefore, one can write

$$\frac{\int H(H_n - H)dH_n}{\sqrt{\int H_n^2 dH_n} + \sqrt{\int H^2 dH_n}} - \frac{\int H(F_nG_n - H)dH_n}{\sqrt{\int F_n^2 G_n^2 dH_n} + \sqrt{\int H^2 dH_n}}$$
$$= \frac{3}{2}\int H(H_n - F_nG_n)dH_n + R_{1n},$$

with $R_{1n} = o_{a.s.}\left(\left|\int H(H_n - H)dH_n\right| + \left|\int H(F_nG_n - H)dH_n\right|\right)$. By Lemma 3.3, $\sqrt{n}R_{1n} = o_{\mathbb{P}}(1)$ and consequently,

$$2n\left(\frac{\int H(H_n - H)\,dH_n}{\sqrt{\int H_n^2\,dH_n} + \sqrt{\int H^2\,dH_n}} - \frac{\int H(F_nG_n - H)\,dH_n}{\sqrt{\int F_n^2 G_n^2\,dH_n} + \sqrt{\int H^2\,dH_n}}\right)^2$$
$$= \frac{9}{2}\left(\sqrt{n}\int H(H_n - F_nG_n)\,dH_n\right)^2 + o_{\mathbb{P}}(1).$$

By Lemma 3.1, it follows that

$$\frac{\sqrt{\int H^2 dH_n}}{\sqrt{\int H_n^2\,dH_n} + \sqrt{\int H^2\,dH_n}} \xrightarrow[n\to\infty]{a.s.} \frac{1}{2},$$
$$\frac{\sqrt{\int H^2 dH_n}}{\sqrt{\int F_n^2 G_n^2\,dH_n} + \sqrt{\int H^2\,dH_n}} \xrightarrow[n\to\infty]{a.s.} \frac{1}{2}.$$

Therefore,

$$\sqrt{\int H^2 dH_n}\Big(\frac{\int (H_n - H)^2 dH_n}{\sqrt{\int H_n^2\,dH_n} + \sqrt{\int H^2\,dH_n}} + \frac{\int (F_nG_n - H)^2 dH_n}{\sqrt{\int F_n^2 G_n^2\,dH_n} + \sqrt{\int H^2\,dH_n}}\Big)$$
$$-\int (H_n - H)(F_nG_n - H)dH_n = \frac{1}{2}\int (H_n - F_nG_n)^2 dH_n + R_{2n},$$

where $R_{2n} = o_{a.s.}\left(\int (H_n - H)^2 dH_n + \int (F_nG_n - H)^2 dH_n\right)$. By Lemma 3.3, $nR_{2n} = o_{\mathbb{P}}(1)$.
Combining the previous results, it follows that

$$nN_n = \frac{9}{2}\left(\sqrt{n}\int H(H_n - F_nG_n)dH_n\right)^2 - \frac{n}{2}\int (H_n - F_nG_n)^2 dH_n + o_{\mathbb{P}}(1). \quad (3.13)$$

The result follows from (3.13), Lemma 3.3, Remark 3.1 and the continuous mapping theorem. □

By construction, T_n is a random variable that takes non-positive values. On the other hand, under the null hypothesis, it holds

$$\int H^2(x,y)dH(x,y) = \int F^2(x)dF(x) \int G^2(x)dG(x) = \frac{1}{9}.$$

So, the Cauchy-Schwarz inequality

$$\begin{aligned}\left(\int H(x,y)U(x,y)dH(x,y)\right)^2 &\leq \int H^2(x,y)dH(x,y) \int U^2(x,y)dH(x,y) \\ &= \frac{1}{9}\int U^2(x,y)dH(x,y).\end{aligned}$$

yields that the limiting distribution is also non-positive. Theorem 3.3 states that the limiting distribution is a random variable that can be represented as the difference of two components. The first component,

$$\frac{81}{2}\left(\int H(x,y)\,U(x,y)\,dH(x,y)\right)^2,$$

has the distribution of the square of a centered univariate normal random variable and is proportional to a χ^2_1 random variable. The second component,

$$\frac{9}{2}\int U^2(x,y)\,dH(x,y),$$

is proportional to $\sum_{k\geq 1}\lambda_k\chi^2_{1,k}$, where $\chi^2_{1,1}, \chi^2_{1,2}, \ldots$ are independent χ^2_1 random variables, and $\{\lambda_k\}_{k\geq 1}$ are the eigenvalues of the integral operator

$$A : L^2(\mathbb{R}^2, H) \to L^2(\mathbb{R}^2, H),$$

defined by

$$Af(x,y) = \int C((x,y),(u,v))\,f(u,v)\,dH(u,v),$$

for $f \in L^2(\mathbb{R}^2, H)$, where C denotes the covariance function of $\mathbb{U}$. The space is defined as

$$L^2(\mathbb{R}^2, H) = \{f : \mathbb{R}^2 \to \mathbb{R} \text{ measurable with } \int f^2(x,y)\,dH(x,y) < \infty\}.$$

Let $\alpha \in (0,1)$ be fixed and consider the statistic T_n to testing H_0. The correspond-

ing test function is defined by

$$\Psi = \begin{cases} 1, & \text{if } T_{n,\text{obs}} \leq q_{n,\alpha}, \\ 0, & \text{otherwise,} \end{cases}$$

where $q_{n,\alpha} = \sup\{t : \mathbb{P}_0(T_n \leq t) \leq \alpha\}$ denotes the α-quantile of the distribution of T_n under H_0, and $T_{n,\text{obs}}$ is the observed value of the statistic. Equivalently, the test rejects H_0 whenever the p-value $p = \mathbb{P}_0(T_n \leq T_{n,\text{obs}})$ is less than or equal to α.

An immediate consequence of Theorems 3.1 and 3.3 is that the test Ψ is consistent. That is, for any fixed alternative (when X does not follow H_0),

$$\mathbb{P}\{\Psi = 1) \to 1 \text{ as } n \to \infty.$$

This property remains valid even if $\mathbb{P}_0$ is replaced by a consistent estimator of the null distribution.

3.2 Goodness-of-fit testing for selected distributions

This section focuses on evaluating the proposed method for gof testing within three specific distribution families and assessing its power against existing tests for finite sample sizes. For benchmarking purposes, the proposed test is compared with the Hoeffding-Blum-Kiefer-Rosenblatt (HBKR) test, whose test statistic is given in (3.4), as well as with two classical tests, namely the Kolmogorov-Smirnov (KS) and Cramér-von Mises (CvM) tests, whose corresponding statistics are given in (3.2) and (3.3), respectively. These tests were selected in order to compare the proposed procedure with methods of different methodological nature. In particular, the HBKR test is also based on an independence-type characterization, making it closely related to the proposed approach, while the KS and CvM tests serve as classical empirical distribution function-based benchmarks.

In general, the null distribution of these test statistics depends on the value of $\boldsymbol{\theta}$. For this reason, their null distribution is approximated by a parametric bootstrap procedure, described as follows:

Algorithm 2 Parametric bootstrap approximation of the null distribution

1: Given the data $X_1, \ldots, X_n$, compute the statistic $S_n = S_n(X_1, \ldots, X_n)$.
2: Compute an estimator $\hat{\boldsymbol{\theta}}_{\boldsymbol{n}} = \hat{\boldsymbol{\theta}}_{\boldsymbol{n}}(X_1, \ldots, X_n)$ of the unknown parameter $\boldsymbol{\theta} \in \Theta$.
3: Generate a bootstrap sample $X_1^*, \ldots, X_n^* \overset{iid}{\sim} F(\cdot; \hat{\boldsymbol{\theta}}_{\boldsymbol{n}})$.
4: Compute the bootstrap statistic $S_n^* = S_n(X_1^*, \ldots, X_n^*)$.
5: Repeat steps 3–4 B times to obtain the empirical distribution of S_n^* conditional on the observed data.
6: The parametric bootstrap estimates $\mathbb{P}_0(S_n \leq t)$ by means of $\mathbb{P}_*(S_n^* \leq t)$ where $\mathbb{P}_*$ denotes the conditional distribution given $X_1, \ldots, X_n$.

The consistency of the parametric bootstrap approximation for KS_n and CvM_n was established in Stute et al., 1993. For the HBKR test statistic, the corresponding result is explicitly stated in Baringhaus and Gaigall, 2015. More precisely, under suitable regularity assumptions, it is shown there that the conditional distribution of the bootstrap statistic and the null distribution of the original statistic, evaluated at the almost sure limit of the parameter estimator, converge to the same limiting distribution. Consequently, if the estimator $\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}$ is strongly consistent for the true parameter value $\boldsymbol{\theta}$ under the null hypothesis, the parametric bootstrap provides a consistent approximation to the null distribution of the statistic. The consistency of $\mathbb{P}_*(T_n^* \leq t)$ as an estimator of the null distribution of T_n can be stated analogously, so its proof is omitted.

Having established the asymptotic framework and described the bootstrap approximation used to obtain critical values, the finite-sample performance of the proposed test is now investigated. To this end, a variety of parametric families were considered as alternative distributions, with details provided in Section 3.3. For each family, several sets of parameter values were selected in order to represent different distributional shapes.

3.2.1 Family of gamma distributions

As a first example, the gof problem for the gamma family is considered, with reference to the independence-type characterization stated in Example 8.

Let $\mathcal{G}$ denote the nonparametric class of absolutely-continuous distributions F concentrated on the positive half-line and having a finite second moment

$$\int_0^\infty x^2 \, dF(x) < \infty. \tag{3.14}$$

Let $X_1, \ldots, X_n$ be a sample of independent copies of a positive random variable X

with distribution $F \in \mathcal{G}$. The aim is to test whether the underlying distribution F belongs to the parametric family of gamma distributions. So, the null hypothesis is given by

$$H_0 : X \sim \Gamma(\alpha, \beta) \quad \text{for some } \alpha > 0 \text{ and } \beta > 0,$$

where the density of $\Gamma(\alpha, \beta)$ is

$$f(x; \alpha, \lambda) = \frac{1}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \beta^{\alpha}, \quad x > 0.$$

The parameter vector $\boldsymbol{\theta} = (\alpha, \beta)$ is estimated using the method of moments. The corresponding estimator is given by $\hat{\boldsymbol{\theta}}_{\boldsymbol{n}} = (\bar{X}_n^2/S_n^2, \bar{X}_n/S_n^2)$, where $\bar{X}_n$ and S_n^2 represent the sample mean and the sample variance, respectively. These estimators are consistent. Indeed, under condition (3.14), the law of large numbers yields $\bar{X}_n \xrightarrow{P} \mathbb{E}(X) = \alpha/\beta$ and $S_n^2 \xrightarrow{P} \mathrm{Var}(X) = \alpha/\beta^2$. Hence, by the continuous mapping theorem $\hat{\alpha}_n \xrightarrow{P} \alpha$ and $\hat{\beta}_n \xrightarrow{P} \beta$.

For the numerical study, the alternatives listed in Section 3.3 were considered under several parameter settings. Table 3.1 presents the empirical sizes and powers for sample sizes $n = 10, 20$ and 50, at the significance level $\alpha = 0.05$. All values are reported as percentages. The simulation study was based on $N = 1000$ Monte Carlo replications, and the bootstrap critical values were obtained from $B = 1000$ bootstrap samples in each replication. The highest power values are highlighted in bold.

Since $N = 1000$, the standard error of an empirical rejection probability is bounded above by $0.5/\sqrt{1000}$. Therefore, on the percentage scale, the corresponding margin $1.96 \cdot 0.5/\sqrt{1000}$ is approximately 3 percentage points. Grey-shaded cells indicate powers that are within this margin of the largest observed power, including the maximum itself. Thus, each grey-shaded cell contains either the maximum power or a value very close to it.

According to the results reported in Table 3.1, the considered tests are generally well calibrated when the shape parameter of the gamma distribution under the null hypothesis is greater than 1, with empirical sizes close to the nominal significance level. An exception is observed for the $\Gamma(0.5, 1)$ distribution, where the classical tests exhibit liberal behavior. To investigate this phenomenon in more detail, an extended simulation study was conducted and the results are presented in Table 3.2. This additional study shows that, for the classical tests, the type I error rate increases as the shape parameter decreases, approaching 1 as the parameter tends to zero. This sensitivity may compromise the reliability of power estimates, particularly for alternatives involving small values of the

estimated shape parameter used in the bootstrap procedure. As a result, the reported power of the classical tests under such conditions should be interpreted with caution. For this reason, the table also reports the theoretical gamma shape parameter $\alpha_0 = (\mathbb{E}[X])^2/\mathrm{Var}(X)$, obtained by matching the first two moments with those of the alternative distribution.

		$n = 10$				$n = 20$				$n = 50$			
Dist.	α_0	$HBKR_n$	T_n	KS_n	CvM_n	$HBKR_n$	T_n	KS_n	CvM_n	$HBKR_n$	T_n	KS_n	CvM_n
$\Gamma(5,1)$		5	5	5	5	4	6	5	5	6	5	4	3
$\Gamma(10,1)$		6	6	5	5	6	5	5	4	6	6	5	5
$\Gamma(2,1)$		4	4	5	6	5	4	4	5	5	5	4	4
$\Gamma(0.5,1)$		4	4	13	11	5	4	10	10	5	5	8	8
$\Gamma(2,2)$		4	5	5	4	6	5	6	6	6	6	4	6
IG$(1,0.5)$	0.5	**19**	13	10	8	**40**	34	25	23	**84**	81	63	58
IG$(0.5,0.5)$	1	**12**	9	7	7	**24**	19	17	17	**58**	53	38	36
IG$(1,2)$	2	**8**	5	6	6	**17**	13	14	14	**33**	27	22	25
LFR$(1,2)$	1.9	6	**9**	7	8	**10**	**10**	7	9	21	**24**	13	15
LFR$(1,4)$	2.16	7	**9**	6	8	10	**14**	8	10	26	**28**	15	20
W$(0.4,3)$	0.1	7	6	**13**	**13**	**13**	10	6	7	**35**	31	24	22
W$(5,1)$	19.06	9	**13**	9	9	15	**18**	12	12	40	**42**	25	30
W$(3,1)$	7.57	9	**11**	8	9	12	**14**	10	10	29	**31**	17	20
W$(5,3)$	19.06	11	**14**	8	8	17	**20**	13	14	43	**46**	26	32
HN(1)	1.75	6	**8**	6	**8**	9	**10**	8	9	18	**20**	15	16
U$(0,1)$	3	22	**28**	17	20	54	**58**	35	46	94	**95**	75	87
U$(1,2)$	27	10	**12**	8	**12**	15	**20**	15	**20**	44	50	34	**53**
U$(0,2)$	3	25	**31**	18	23	54	**59**	32	46	95	**96**	75	90
B$(2,2)$	5	14	**18**	9	9	28	**31**	17	23	67	**70**	41	52
B$(0.5,0.5)$	2	41	**47**	36	45	83	**85**	68	81	**100**	**100**	99	99
B$(5,1)$	35	39	**45**	28	33	76	**80**	54	66	**100**	**100**	92	98
P$(3,20)$	15	36	**44**	26	30	75	**77**	48	61	**99**	**99**	91	98
LG$(3,6)$	7.67	**20**	12	16	19	**44**	34	33	41	**89**	86	72	83
LG$(2,3)$	1.34	**31**	20	23	27	**68**	58	54	60	**98**	97	92	95
Gomp$(1,0.25)$	3.38	16	**20**	13	14	32	**36**	18	23	71	**74**	47	59
Gomp$(1,0.5)$	2.58	10	**13**	11	12	22	**25**	15	17	51	**53**	32	41

Table 3.1: Testing goodness-of-fit to the gamma distribution: Empirical powers (in percentages)

Table 3.1 shows that no single test achieves the highest power across all alternatives, meaning there is no uniformly superior test. This aligns with theoretical results from Janssen, 2000, which state that the global power function of any nonparametric test remains flat on balls of alternatives, except for those from a finite-dimensional subspace. In most cases, the tests based on independence character-

ization demonstrate greater power than the classical tests. The independence-based tests examined here exhibit comparable power in some cases, while in other situations notable differences arise. For instance, the HBKR test outperforms the newly proposed test under log-gamma and inverse Gaussian alternatives, whereas the proposed test performs better for alternatives with finite support, such as the beta, uniform and power function distributions.

	$n = 10$				$n = 20$				$n = 50$			
Dist.	$HBKR_n$	T_n	KS_n	CvM_n	$HBKR_n$	T_n	KS_n	CvM_n	$HBKR_n$	T_n	KS_n	CvM_n
$\Gamma(0.2, 1)$	3	3	25	26	4	4	17	19	4	4	10	11
$\Gamma(0.4, 1)$	4	4	15	14	4	4	9	10	4	4	9	9
$\Gamma(0.6, 1)$	4	4	8	8	4	4	8	9	4	4	8	8
$\Gamma(0.8, 1)$	4	4	6	6	5	4	7	6	4	4	5	5

Table 3.2: Additional simulations: Empirical sizes (in percentages)

3.2.2 Family of inverse Gaussian distributions

The second testing problem considered in this chapter concerns the family of inverse Gaussian distributions, whose independence characterization was presented in Example 9.

Similar to the preceding, let $\mathcal{G}$ denote the nonparametric class of absolutely-continuous distributions F concentrated on the positive half-line such that

$$\int_0^\infty x\, dF(x) < \infty \text{ and } \int_0^\infty \frac{1}{x}\, dF(x) < \infty.$$

Let $X_1, \ldots, X_n$ be a sample of independent copies of a positive random variable X with distribution $F \in \mathcal{G}$. The aim is to test whether the underlying distribution F belongs to the parametric family of inverse Gaussian distributions. So, the null hypothesis is

$$H_0 : X \sim IG(\mu, \lambda) \quad \text{for some } \mu > 0 \text{ and } \lambda > 0,$$

where the density of the inverse Gaussian distribution is given by

$$f(x; \mu, \lambda) = \left(\frac{\lambda}{2\pi x^3}\right)^{1/2} \exp\left\{-\frac{\lambda(x-\mu)^2}{2\mu^2 x}\right\}, \quad x > 0.$$

The parameter vector $\boldsymbol{\theta} = (\mu, \lambda)$ is estimated with its maximum likelihood estimator $\hat{\boldsymbol{\theta}}_{\boldsymbol{n}} = \left(\bar{X}_n, \left(\frac{1}{n}\sum_i \frac{1}{X_i} - \frac{1}{\bar{X}_n}\right)^{-1}\right)$. The alternatives are identical to those mentioned in the previous subsection. Table 3.3 presents the empirical sizes and powers (expressed as percentages) for different sample sizes $n = 10, 20, 50$ at a significance level of $\alpha = 0.05$. The results are based on $N = 1000$ simulation replications. The quantiles of the bootstrap distributions (critical values) are obtained by Monte Carlo simulation, each based on $B = 1000$ bootstrap samples. The meaning of the bold numbers and the grey cells is the same as in the previous setting.

As shown in Table 3.3, all tests are generally well calibrated. Among the classical tests, the CvM test outperforms the KS test, while no clear dominance is observed between the characterization-based tests and CvM. Although CvM appears to be the most effective overall for larger sample sizes, the performance of T_n remains competitive for smaller samples, outperforming CvM for a considerable number of alternatives. When comparing characterization-based tests, T_n performs better than HBKR for small samples.

3.2.3 Standard Cauchy distribution

As the last case considered in this chapter, the gof problem for the standard Cauchy distribution is examined using the characterization stated in Example 10.

Let $X_1, \ldots, X_n$ be a sample of independent copies of a real-valued random variable X. The null hypothesis is

$$H_0 : X \sim \mathcal{C}(0, 1),$$

where the density of the Cauchy $\mathcal{C}(\mu, \sigma)$ distribution is

$$f(x; \mu, \sigma) = \frac{1}{\pi\sigma}\left[1 + \left(\frac{x-\mu}{\sigma}\right)^2\right]^{-1}, \quad x \in \mathbb{R}.$$

Since the null hypothesis is simple, critical values can be obtained using the standard Monte Carlo method without bootstrapping. Details of the alternative distributions are provided in Section 3.3. Table 3.4 presents the results obtained for sample sizes $n = 10, 20$, and 50, at the significance level $\alpha = 0.05$, based on $N = 10000$ replications. The interpretation of the bold numbers and grey cells remains the same as in the previous settings, with the only difference being that in this case, the grey cells represent those with a power that is at most

	$n=10$				$n=20$				$n=50$			
Dist.	$HBKR_n$	T_n	KS_n	CvM_n	$HBKR_n$	T_n	KS_n	CvM_n	$HBKR_n$	T_n	KS_n	CvM_n
IG(0.5, 0.5)	6	6	7	8	4	4	7	6	4	4	5	5
IG(0.5, 10)	5	5	7	7	4	5	5	5	5	5	4	5
IG(1, 1)	4	6	5	5	6	5	5	5	4	5	4	3
IG(10, 0.5)	4	4	5	5	5	4	7	6	6	6	5	6
GIG(10, 1, 5)	6	**8**	7	7	7	**9**	8	8	10	**12**	9	10
GIG(10, 10, 5)	4	**6**	5	4	7	**8**	6	5	9	**10**	7	9
GIG(5, 0.1, 5)	6	**9**	8	8	10	**14**	12	13	19	**21**	17	20
$\Gamma(2,1)$	7	12	17	**19**	18	20	30	**31**	44	41	55	**63**
$\Gamma(2,2)$	10	15	16	**18**	17	21	28	**33**	40	40	52	**61**
$\Gamma(5,1)$	6	**10**	8	8	10	**12**	**12**	**12**	18	19	19	**23**
$\Gamma(10,1)$	4	**7**	5	5	6	**8**	6	7	10	**12**	9	10
LFR(1, 2)	19	27	36	**40**	49	49	63	**69**	89	85	94	**96**
LFR(1, 4)	15	25	34	**40**	47	50	62	**69**	86	83	92	**95**
W(5, 1)	11	**18**	14	15	24	**27**	22	**27**	**61**	55	48	59
W(3, 1)	12	**21**	17	19	27	**30**	25	29	60	56	56	**63**
W(5, 3)	10	**16**	11	13	23	**27**	22	**27**	**61**	58	50	**61**
HN(1)	6	**9**	6	7	47	52	64	**69**	90	87	95	**98**
U(0, 1)	43	53	49	**56**	85	82	81	**86**	**100**	99	99	**100**
U(1, 2)	9	**18**	7	9	20	**31**	13	17	53	**60**	35	52
U(0, 2)	44	54	49	**56**	85	86	80	**87**	**100**	99	99	**100**
SE(1, 1)	**16**	6	10	12	**30**	18	18	23	**70**	55	43	57
B(2, 2)	27	**33**	24	29	54	50	51	**58**	92	82	86	**94**
B(0.5, 0.5)	73	80	78	**82**	**98**	**98**	**98**	**98**	**100**	**100**	**100**	**100**
B(5, 1)	35	**46**	31	38	**82**	73	59	74	**99**	96	97	**99**
B(2, 5)	13	**21**	19	**21**	31	32	33	**39**	66	59	68	**77**
P(3, 20)	36	**46**	32	39	**82**	73	63	75	**99**	96	97	**99**
LG(3, 6)	**19**	5	13	15	**36**	12	25	32	**75**	50	58	70
LG(2, 3)	**25**	6	18	22	**53**	23	40	48	**89**	70	73	87
Gomp(1, 0.25)	27	36	40	**43**	66	66	70	**77**	98	95	97	**99**
Gomp(1, 0.5)	22	33	41	**46**	61	62	69	**76**	96	92	**98**	**98**

Table 3.3: Testing goodness-of-fit to the inverse Gaussian distribution: Empirical powers (in percentages)

$1.96 \times 0.5/\sqrt{10000} = 0.01$ (approximately 1%) below the highest value (including the highest itself).

Table 3.4 illustrates that for sample sizes of 10 and 20, the classical tests have extremely low powers against any unimodal alternatives that are symmetric and sharply concentrated around zero, which are the most plausible alternatives to the standard Cauchy distribution. In contrast, characterization-based

tests demonstrate significantly higher power in these scenarios. For a sample size of 50, the powers of classical tests are better, yet they remain lower than those of independence-based tests. By contrast, classical tests are better for detecting shifts or substantially different shapes, such as uniform. For such alternatives, T_n performs reasonably well, while HBKR has extremely low power for small sample sizes. These findings strongly support the use of the novel test for testing the null standard Cauchy distribution.

	$n = 10$				$n = 20$				$n = 50$			
Dist.	$HBKR_n$	T_n	KS_n	CvM_n	$HBKR_n$	T_n	KS_n	CvM_n	$HBKR_n$	T_n	KS_n	CvM_n
C(0, 1)	5	5	5	4	5	6	6	6	6	6	6	6
C(0.5, 1)	3	4	**16**	15	3	12	**27**	**27**	36	54	**58**	**58**
C(1.5, 1)	4	14	**77**	74	14	65	**95**	92	99	**100**	**100**	**100**
C(0, 0.5)	**41**	38	5	3	**64**	57	10	7	**94**	91	40	40
C(0, 1.5)	1	3	**9**	**9**	1	5	**12**	10	11	**23**	17	16
t_3	**22**	20	5	4	**39**	37	4	4	**67**	62	8	9
t_5	**29**	27	4	3	**48**	45	4	4	**86**	82	7	7
N(0, 1)	50	**51**	6	5	**72**	71	5	5	**98**	**98**	29	35
N(0.5, 1)	23	**24**	20	19	44	**65**	46	50	98	**100**	92	95
N(0, 1.5)	6	**7**	6	6	11	15	6	5	14	**16**	6	7
N(0, 2)	2	8	**9**	7	3	**15**	10	7	24	**37**	18	12
Lo(0, 1)	4	5	**6**	5	5	**11**	6	5	**100**	**100**	**100**	**100**
Lo(0, 2)	0	11	**20**	16	9	**34**	31	24	82	**94**	70	64
La(0, 1)	**27**	26	3	2	**49**	47	5	3	**100**	**100**	**100**	**100**
La(0.5, 1)	13	16	19	**20**	28	**50**	41	42	88	**94**	81	84
La(0, 1.5)	6	**7**	4	5	5	**6**	5	4	10	**12**	6	6
La(0, 2)	1	4	**10**	9	2	**8**	**8**	7	10	**12**	6	6
Gu(0, 1)	**22**	15	12	13	38	**45**	28	35	86	**92**	89	84
NC(0.3)	**13**	12	4	4	**20**	16	4	3	**42**	33	7	6
NC(0.5)	**24**	23	5	4	**37**	31	4	3	**69**	61	12	9
Tu(1)	**10**	9	3	4	**10**	8	4	5	**16**	12	6	5
U(−10, 10)	1	26	**5**	42	21	74	86	**93**	**100**	**100**	**100**	**100**
St(1.5, 0)	**8**	7	4	4	10	**12**	5	4	14	**15**	5	5
St(1.5, 0.5)	9	**10**	6	6	9	**13**	7	7	15	**17**	11	12

Table 3.4: Testing goodness-of-fit to the Cauchy distribution: Empirical powers (in percentages)

3.3 The list of alternative distributions

- The generalised inverse Gaussian distribution GIG(α, β, θ), with $\alpha, \beta > 0$ and $\theta \in \mathbb{R}$, has density

$$\frac{(\alpha/\beta)^{\theta/2}}{2K_\theta(\sqrt{\alpha\beta})} x^{\theta-1} e^{-(\alpha x+\beta/x)/2}, \quad x > 0, \tag{3.15}$$

  where K_θ is modified Bessel function of the third kind with index θ.

- The linear failure rate distribution LFR(μ, θ), with $\mu, \theta > 0$, has density

$$(\mu + \theta x) e^{-\mu x - \frac{1}{2}\theta x^2}, \quad x > 0. \tag{3.16}$$

- The Weibull distribution W(μ, θ), with $\mu, \theta > 0$, has density

$$\frac{\mu x^{\mu-1}}{\theta^\mu} e^{-(\frac{x}{\theta})^\mu}, \quad x > 0. \tag{3.17}$$

- The half-normal distribution HN(θ), with $\theta > 0$, has density

$$\frac{\sqrt{2}}{\theta\sqrt{\pi}} e^{-\frac{x^2}{2\theta^2}}, \quad x > 0. \tag{3.18}$$

- The uniform distribution U(θ_1, θ_2), with $-\infty < \theta_1 < \theta_2 < +\infty$, has density

$$\frac{1}{\theta_2 - \theta_1}, \quad \theta_1 < x < \theta_2. \tag{3.19}$$

- The shifted exponential distribution SE(μ, θ), with $\mu \in \mathbb{R}$ and $\theta > 0$, has density

$$\theta e^{-\theta(x-\mu)}, \quad x > \mu. \tag{3.20}$$

- The beta distribution B(θ_1, θ_2), with $\theta_1, \theta_2 > 0$, has density

$$\frac{\Gamma(\theta_1 + \theta_2)}{\Gamma(\theta_1)\Gamma(\theta_2)} x^{\theta_1-1}(1-x)^{\theta_2-1}, \quad 0 < x < 1. \tag{3.21}$$

- The power function distribution P(μ, θ), with $\mu, \theta > 0$, has density

$$\mu\left(\frac{x}{\theta}\right)^\mu \frac{1}{x}, \quad 0 < x < \theta. \tag{3.22}$$

- The loggamma distribution $\mathrm{LG}(\mu, \theta)$, with $\mu, \theta > 0$, has density

$$\frac{\theta^{\mu}}{\Gamma(\mu)} \frac{(\log(x))^{\mu-1}}{x^{\theta+1}}, \quad x > 1. \tag{3.23}$$

- The Gompertz distribution $\mathrm{Gomp}(\mu, \theta)$, with $\mu, \theta > 0$, has density

$$\theta e^{\mu x} e^{-\frac{\theta(e^{\mu x}-1)}{\mu}}, \quad x > 0. \tag{3.24}$$

- The Cauchy distribution $\mathrm{C}(\mu, \sigma)$, with $\mu \in \mathbb{R}$ and $\sigma > 0$, has density

$$\frac{1}{\pi} \frac{\sigma}{(x-\mu)^2 + \sigma^2}, \quad x \in \mathbb{R}. \tag{3.25}$$

- The normal distribution $\mathrm{N}(\mu, \sigma^2)$, with $\mu \in \mathbb{R}$ and $\sigma > 0$, has density

$$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x \in \mathbb{R}. \tag{3.26}$$

- The Student t distribution t_{ν}, with $\nu \in \mathbb{Z}^+$, has density

$$\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}, \quad x \in \mathbb{R}. \tag{3.27}$$

- The logistic distribution $\mathrm{Lo}(\mu, \sigma)$, with $\mu \in \mathbb{R}$ and $\sigma > 0$, has density

$$\frac{e^{-\frac{(x-\mu)}{\sigma}}}{\sigma\left(1 + e^{-\frac{(x-\mu)}{\sigma}}\right)^2}, \quad x \in \mathbb{R}. \tag{3.28}$$

- The Laplace distribution $\mathrm{La}(\mu, \sigma)$, with $\mu \in \mathbb{R}$ and $\sigma > 0$, has density

$$\frac{1}{2\sigma} e^{-\frac{|x-\mu|}{\sigma}}, \quad x \in \mathbb{R}. \tag{3.29}$$

- The Gumbel distribution $\mathrm{Gu}(\mu, \sigma)$, with $\mu \in \mathbb{R}$ and $\sigma > 0$, has density

$$\frac{1}{\sigma} e^{-\left(\frac{x-\mu}{\sigma} + e^{-\frac{(x-\mu)}{\sigma}}\right)}, \quad x \in \mathbb{R}. \tag{3.30}$$

- The standard Gaussian-Cauchy mixture distribution $\mathrm{NC}(\theta)$, with mixture

parameter $\theta \in (0,1)$, has density

$$\theta \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} + (1-\theta) \frac{1}{\pi(1+x^2)}, \quad x \in \mathbb{R}. \tag{3.31}$$

- The Tukey h-distribution Tu(h), with $h \geq 0$, is defined as the distribution of a transformation of a standard normal random variable Z by

$$Z e^{\frac{hZ^2}{2}}. \tag{3.32}$$

- The stable distribution St(α, β), with $\alpha \in (0,2]$ and $\beta \in [-1,1]$, has characteristic function

$$g(t) = \begin{cases} \exp\left(-|t|^\alpha \left(1 - i\beta \operatorname{sgn}(t) \tan\left(\frac{\alpha\pi}{2}\right)\right)\right), & \alpha \neq 1, \\ \exp\left(-|t| \left(1 + \frac{2}{\pi} i\beta \operatorname{sgn}(t) \log|t|\right)\right), & \alpha = 1, \end{cases} \qquad t \in \mathbb{R}.$$

Chapter 4

A goodness-of-fit test for the wrapped normal distribution

Circular data consist of measurements expressed as directions, angles, or positions on a periodic scale. Unlike linear data, circular measurements have no natural beginning or end, since the endpoints of the scale coincide. This type of data naturally occurs in many applied fields, for example, when analyzing directions of sea currents in oceanography, orientations of plants toward sunlight in environmental science, or the times of day at which traffic accidents occur in transportation research. In such cases, the observations are treated more appropriately as points on a circle than as values on the real line.

The analysis of circular data requires methods that account for their specific geometric structure. In particular, circular observations are invariant under the choice of origin, and the distance between two points is determined by the shorter arc connecting them on the circle. These features distinguish circular data fundamentally from linear data and make standard statistical techniques developed for Euclidean settings generally unsuitable without appropriate modification. Consequently, the study of probabilistic models and statistical procedures tailored to circular observations forms an important part of directional statistics.

One of the most common ways to obtain circular distributions is to wrap a real-valued random variable around the circle by means of the modulo 2π transformation. This transformation maps the real line onto the circumference of a unit circle, resulting in the accumulation of probabilities at points where the line wraps around the circle multiple times. Among the models developed for circular data, the wrapped normal distribution occupies an important place as the circular analog of the normal distribution. It has been used in numerous applied contexts where observations are naturally represented on a circle. For example, it has been used to model orientation and heading in periodic state spaces, particularly in bearing-only sensor scheduling and constrained object tracking on circular man-

ifolds (see, e.g., Gilitschenski et al. 2013; Kurz et al. 2013). In oceanography, the wrapped normal distribution is widely used to model wave directions (see, e.g., Jona Lasinio et al. 2012), as the underlying variability tends to follow a symmetric and smooth pattern. Furthermore, in cellular chronobiology, the wrapped normal distribution is applied to model cellular circadian phases, enabling quantification of population synchrony and amplitude variability (see, e.g., St. John et al. 2014). These examples illustrate the practical relevance of the wrapped normal distribution as a flexible and interpretable model for circular data.

Let Θ be a random variable whose pdf is given by

$$f(\theta) = \frac{1}{\sigma\sqrt{2\pi}} \sum_{m=-\infty}^{\infty} exp\Big[\frac{-(\theta - \mu + 2\pi m)^2}{2\sigma^2}\Big],\ -\pi \leq \theta, \mu < \pi,\ \sigma > 0. \qquad (4.1)$$

Then Θ is said to follow the wrapped normal distribution, denoted by $\Theta \sim \mathcal{WN}(\mu, \sigma)$. The main goal of this chapter is to propose a new gof test for the wrapped normal family. More precisely, let Θ be a circular random variable with unknown distribution, and let $\Theta_1, \ldots, \Theta_n$ denote i.i.d. observations from the distribution of Θ. The null hypothesis is

$$H_0 : \Theta \sim \mathcal{WN}(\mu, \sigma) \ \text{ for some } \ -\pi \leq \mu < \pi \ \text{ and } \ \sigma > 0,$$

against the alternative:

$$H_1 : \Theta \not\sim \mathcal{WN}(\mu, \sigma) \ \text{ for all } \ -\pi \leq \mu < \pi \ \text{ and } \sigma > 0.$$

One common approach to gof testing involves comparing the edf of the observed data with the parametric cdf specified under the null hypothesis. The distance between these two functions serves as the basis for various test statistics. For example, the Kuiper and Watson tests (Kuiper 1960; Watson 1961), circular adaptations of the Kolmogorov-Smirnov and Cramér-von Mises tests, are obtained by using discrepancy measures adapted to data on the circle rather than on the real line. Another important approach to gof testing relies on characteristic functions as a basis for constructing test statistics (see, e.g., Jammalamadaka et al. 2019; García-Portugués et al. 2024). Although gof tests based on independence characterizations for real-valued distributions have proven effective in detecting subtle departures from the null hypothesis, often outperforming classical methods (see, e.g., Milošević and Obradović 2016; Baringhaus and Gaigall 2015; Halaj et al. 2024), their direct application to circular data remains problematic. The main difficulty lies in the absence of a natural origin and orientation on the unit circle.

Consequently, procedures developed for linear data generally do not possess the rotation and reflection invariance required for meaningful inference in the circular setting. This leaves an important open direction in gof testing, namely the development of independence characterization-based procedures that are intrinsically suited to circular distributions.

This chapter seeks to fill the identified gap by developing a new gof test for the wrapped normal distribution, derived from an independence characterization, demonstrating its potential for circular data analysis. The basis for the proposed test is given by the following theorem, stated in Rukhin, 1972.

Theorem 4.1. *Let Θ_1, Θ_2 be i.i.d. circular random variables taking values in $[0, 2\pi)$. Assume that their common distribution does not coincide with any of its nontrivial circular shifts and that the distribution of $\Theta_1 +_{2\pi} \Theta_2$ is absolutely continuous. Then $\Theta_1 +_{2\pi} \Theta_2$ and $\Theta_1 -_{2\pi} \Theta_2$ are independent only if both random variables have wrapped normal distribution (defined on [0,2π)).*

In this theorem, angles are defined on the interval $[0, 2\pi)$, with their sum and difference are taken modulo 2π, i.e.,

$$\begin{aligned}\Theta_1 +_{2\pi} \Theta_2 &= (\Theta_1 + \Theta_2) \bmod 2\pi,\\ \Theta_1 -_{2\pi} \Theta_2 &= (\Theta_1 - \Theta_2) \bmod 2\pi.\end{aligned}$$

Since circular observations in this chapter are represented on the interval $[-\pi, \pi)$, the corresponding operations are written in the following equivalent form. For two circular random variables Θ_1 and Θ_2 taking values in $[-\pi, \pi)$, the operations $+_\pi$ and $-_\pi$ are expressed as

$$\Theta_1 +_\pi \Theta_2 = \big((\Theta_1 + \Theta_2 + \pi) \bmod 2\pi\big) - \pi, \tag{4.2}$$

$$\Theta_1 -_\pi \Theta_2 = \big((\Theta_1 - \Theta_2 + \pi) \bmod 2\pi\big) - \pi. \tag{4.3}$$

Thus, both the circular sum and the circular difference are again represented in the principal interval $[-\pi, \pi)$. It is important to note that Theorem 4.1 holds when these modifications are applied. The assumption that the common distribution differs from all of its nontrivial circular shifts rules out the circular uniform distribution, which remains unchanged under any circular shift. Before introducing the test statistic, the relevant characterization is formulated and proved in the framework used throughout this chapter. For this purpose, let

$$\varphi_\Theta(k) := \mathbb{E}\big(e^{ik\Theta}\big),\ k \in \mathbb{Z},$$

denote the circular characteristic function of a circular random variable Θ. The argument of the circular characteristic function is restricted to the integers because angles that differ by an integer multiple of 2π represent the same point on the circle. Indeed, for a real argument t, the mapping $\theta \mapsto e^{it\theta}$ is well defined on the circle only if

$$e^{it(\theta+2\pi)} = e^{it\theta}$$

for every θ. This condition is equivalent to

$$e^{2\pi it} = 1,$$

which is satisfied iff $t \in \mathbb{Z}$. In particular, the circular characteristic function has the following properties:

$$\varphi_\Theta(0) = 1, \quad |\varphi_\Theta(k)| \leq 1, \quad \varphi_\Theta(-k) = \overline{\varphi_\Theta(k)}, \quad k \in \mathbb{Z}.$$

Moreover, if Θ_1 and Θ_2 are independent circular random variables, then

$$\varphi_{\Theta_1 +_\pi \Theta_2}(k) = \varphi_{\Theta_1}(k)\varphi_{\Theta_2}(k), \quad k \in \mathbb{Z}.$$

Finally, the sequence $\{\varphi_\Theta(k) : k \in \mathbb{Z}\}$ uniquely determines the distribution of Θ. For these and related properties, see, for example, Mardia and Jupp, 2000; Jammalamadaka and SenGupta, 2001.

Next, the characterization used in the construction of the proposed test is presented and proved.

Characterization 4.1. *Let Θ_1 and Θ_2 be independent copies of a nondegenerate circular random variable Θ taking values in $[-\pi, \pi)$ such that $\varphi_\Theta(k) \neq 0$, $k \in \mathbb{Z}$. Then $\Theta_1 +_\pi \Theta_2$ and $\Theta_1 -_\pi \Theta_2$ are independent iff Θ follows a wrapped normal distribution defined in* (4.1).

Proof. By the definitions of $+_\pi$ and $-_\pi$, for every $k \in \mathbb{Z}$,

$$e^{ik(\Theta_1 +_\pi \Theta_2)} = e^{ik(\Theta_1+\Theta_2)} \text{ and } e^{ik(\Theta_1 -_\pi \Theta_2)} = e^{ik(\Theta_1-\Theta_2)}, \tag{4.4}$$

since the modular correction is always an integer multiple of 2π.

Assume first that $\Theta_1 +_\pi \Theta_2$ and $\Theta_1 -_\pi \Theta_2$ are independent. Then, for all $k, l \in \mathbb{Z}$, it holds

$$\varphi_{\Theta_1 +_\pi \Theta_2,\, \Theta_1 -_\pi \Theta_2}(k, l) = \varphi_{\Theta_1 +_\pi \Theta_2}(k)\, \varphi_{\Theta_1 -_\pi \Theta_2}(l).$$

Using the relations (4.4) and the fact that Θ_1, Θ_2 are i.i.d., it follows that

$$\begin{aligned}\varphi_{\Theta_1+_\pi\Theta_2,\,\Theta_1-_\pi\Theta_2}(k,l) &= \mathbb{E}\left[e^{ik(\Theta_1+\Theta_2)}e^{il(\Theta_1-\Theta_2)}\right]\\ &= \mathbb{E}\left[e^{i(k+l)\Theta_1}e^{i(k-l)\Theta_2}\right]\\ &= \varphi_\Theta(k+l)\varphi_\Theta(k-l),\end{aligned}$$

whereas

$$\begin{aligned}\varphi_{\Theta_1+_\pi\Theta_2}(k)\,\varphi_{\Theta_1-_\pi\Theta_2}(l) &= \mathbb{E}\left[e^{ik(\Theta_1+\Theta_2)}\right]\mathbb{E}\left[e^{il(\Theta_1-\Theta_2)}\right]\\ &= \varphi_\Theta(k)^2\,\varphi_\Theta(l)\varphi_\Theta(-l)\\ &= \varphi_\Theta(k)^2\,|\varphi_\Theta(l)|^2.\end{aligned}$$

Hence,

$$\varphi_\Theta(k+l)\varphi_\Theta(k-l) = \varphi_\Theta(k)^2\,|\varphi_\Theta(l)|^2,\ \ k,l\in\mathbb{Z}. \tag{4.5}$$

Write the circular characteristic function in its polar form as

$$\varphi_\Theta(k) = r_k e^{i\psi_k},\quad k\in\mathbb{Z},$$

where $r_k := |\varphi_\Theta(k)|$ and $\psi_k := \arg(\varphi_\Theta(k))$. Then (4.5) yields

$$r_{k+l}r_{k-l}e^{i(\psi_{k+l}+\psi_{k-l})} = r_k^2 r_l^2 e^{2i\psi_k},\quad k,l\in\mathbb{Z}.$$

By comparing moduli, one obtains

$$r_{k+l}r_{k-l} = r_k^2 r_l^2,\quad k,l\in\mathbb{Z}. \tag{4.6}$$

Let $a_k := -\log r_k,\ k\in\mathbb{Z}$. Then (4.6) becomes

$$a_{k+l} + a_{k-l} = 2a_k + 2a_l,\quad k,l\in\mathbb{Z}.$$

Since $r_0 = 1$, it follows that $a_0 = 0$. Taking into account that the circular characteristic function satisfies the Hermitian relation

$$\varphi_\Theta(-k) = \overline{\varphi_\Theta(k)},\ \ k\in\mathbb{Z}, \tag{4.7}$$

it follows that $r_{-k} = r_k$, and hence $a_{-k} = a_k$. Therefore, it is sufficient to determine

a_k for $k \in \mathbb{N}$. Taking $l = 1$ and arguing by induction, one obtains

$$a_k = k^2 a_1, \quad k \in \mathbb{N},$$

and consequently

$$r_k = r_1^{k^2}, \quad k \in \mathbb{N}.$$

Analogously, comparison of the arguments gives

$$\psi_{k+l} + \psi_{k-l} \equiv 2\psi_k \pmod{2\pi}, \ \ k, l \in \mathbb{Z}.$$

Taking $l = 1$ and using $\psi_0 = 0$, an induction argument yields

$$\psi_k \equiv k\psi_1 \pmod{2\pi}, \ \ k \in \mathbb{N}.$$

Moreover, since the circular characteristic function satisfies the Hermitian relation (4.7), the corresponding arguments satisfy

$$\psi_{-k} \equiv -\psi_k \pmod{2\pi}.$$

Therefore, the same representation extends to all $k \in \mathbb{Z}$, that is,

$$\psi_k \equiv k\psi_1 \pmod{2\pi}, \ \ k \in \mathbb{Z}.$$

Therefore,

$$\varphi_\Theta(k) = r_1^{k^2} e^{ik\psi_1} = \exp\left(-\frac{\sigma^2 k^2}{2} + ik\mu\right), \quad k \in \mathbb{Z},$$

where $\mu := \psi_1$ and $\sigma^2 := -2\log r_1$. This is the characteristic function of the wrapped normal distribution.

For the converse, assume that Θ_1 and Θ_2 are i.i.d. wrapped normal random variables with characteristic function

$$\varphi_\Theta(k) = \exp\left(-\frac{\sigma^2 k^2}{2} + ik\mu\right), \quad k \in \mathbb{Z}.$$

Then, for all $k, l \in \mathbb{Z}$,

$$\begin{aligned}
\varphi_{\Theta_1 +_\pi \Theta_2,\, \Theta_1 -_\pi \Theta_2}(k, l) &= \varphi_\Theta(k+l)\varphi_\Theta(k-l) \\
&= \exp\left(-\frac{\sigma^2(k+l)^2}{2} + i(k+l)\mu\right) \exp\left(-\frac{\sigma^2(k-l)^2}{2} + i(k-l)\mu\right) \\
&= \exp\left(-\sigma^2 k^2 + 2ik\mu\right) \exp\left(-\sigma^2 l^2\right).
\end{aligned}$$

On the other hand,

$$\begin{aligned}\varphi_{\Theta_1+_\pi\Theta_2}(k)\,\varphi_{\Theta_1-_\pi\Theta_2}(l) &= \varphi_\Theta(k)^2\,|\varphi_\Theta(l)|^2 \\ &= \exp\Big(-\sigma^2k^2+2ik\mu\Big)\exp\Big(-\sigma^2l^2\Big).\end{aligned}$$

Thus,

$$\varphi_{\Theta_1+_\pi\Theta_2,\,\Theta_1-_\pi\Theta_2}(k,l) = \varphi_{\Theta_1+_\pi\Theta_2}(k)\,\varphi_{\Theta_1-_\pi\Theta_2}(l) \quad \text{for all } k,l\in\mathbb{Z},$$

which proves that $\Theta_1 +_\pi \Theta_2$ and $\Theta_1 -_\pi \Theta_2$ are independent. □

4.1 Construction of the novel test statistic

According to the characterization stated above, under the assumption that the circular characteristic function of Θ is non-vanishing on $\mathbb{Z}$, a circular random variable Θ has the wrapped normal distribution $\mathcal{WN}(\mu,\sigma)$, for some $\mu\in[-\pi,\pi)$ and $\sigma>0$, iff

$$\varphi(k,l)=\varphi(k,0)\varphi(0,l),\ \ k,l\in\mathbb{Z},$$

where

$$\varphi(k,l)=\mathbb{E}\left(e^{ik(\Theta_1+_\pi\Theta_2)+il(\Theta_1-_\pi\Theta_2)}\right)$$

denotes the joint characteristic function of the random vector $(\Theta_1 +_\pi \Theta_2, \Theta_1 -_\pi \Theta_2)$, where Θ_1 and Θ_2 are i.i.d. copies of Θ. This motivates a test for H_0 based on the discrepancy between the joint empirical characteristic function and the product of the corresponding marginal empirical characteristic functions.

Let $\Theta_1,\ldots,\Theta_n$ be an i.i.d. sample from the distribution of circular random variable Θ. The joint V-empirical characteristic function corresponding to the random vector formed by the circular sum and circular difference is defined by

$$\varphi_n(k,l)=\frac{1}{n^2}\sum_{j=1}^{n}\sum_{r=1}^{n}e^{ik(\Theta_j+_\pi\Theta_r)+il(\Theta_j-_\pi\Theta_r)},\ \ k,l\in\mathbb{Z}.$$

Therefore, the proposed test statistic is defined as follows:

$$S_{n,\omega}=n\sum_{k=-\infty}^{\infty}\sum_{l=-\infty}^{\infty}|\varphi_n(k,l)-\varphi_n(k,0)\varphi_n(0,l)|^2\omega(k,l),$$

where $|\cdot|$ is a modulus of a complex number and $\omega:\mathbb{Z}^2\to(0,\infty)$ denotes a

positive weight function that satisfies the condition

$$\sum_{k=-\infty}^{\infty}\sum_{l=-\infty}^{\infty}\omega(k,l)<\infty.$$

Consider the separable Hilbert space $l^2_\omega(\mathbb{Z}^2)$ of all sequences $a:\mathbb{Z}^2\to\mathbb{C}$ satisfying

$$\sum_{k=-\infty}^{\infty}\sum_{l=-\infty}^{\infty}|a_{kl}|^2\omega(k,l)<\infty$$

with the scalar product and the norm defined by

$$\langle a,b\rangle_{l^2_\omega(\mathbb{Z}^2)}=\sum_{k=-\infty}^{\infty}\sum_{l=-\infty}^{\infty}a_{kl}\overline{b}_{kl}\omega(k,l)\ \text{ and }\ \|a\|_{l^2_\omega(\mathbb{Z}^2)}=\sqrt{\langle a,a\rangle_{l^2_\omega(\mathbb{Z}^2)}},$$

respectively. With this in mind, the test statistic can be expressed as

$$S_{n,\omega}=||D_n||^2_{l^2_\omega(\mathbb{Z}^2)}$$

where $D_n(k,l)=\sqrt{n}(\varphi_n(k,l)-\varphi_n(k,0)\varphi_n(0,l))$. Under the null hypothesis, the test statistic is expected to take values close to zero. Substantial deviations of $S_{n,\omega}$ from zero indicate a departure from the null hypothesis, so large values of $S_{n,\omega}$ provide evidence against it.

Having introduced the test statistic, the next step is to investigate its large-sample behavior under the null hypothesis.

4.2 Asymptotic properties

Theorem 4.2 (Halaj, 2025)**.** *Under the null hypothesis, if the weight function ω is positive and*

$$\sum_{k=-\infty}^{\infty}\sum_{l=-\infty}^{\infty}\omega(k,l)<\infty, \tag{4.8}$$

then there exists a centered Gaussian element $\mathcal{G}\in l^2_\omega(\mathbb{Z}^2)$ with covariance kernel

$$K_{\mathcal{G}}((k_1,l_1),(k_2,l_2);\sigma)=\mathbb{E}\Gamma(\Theta;k_1,l_1,\sigma)\overline{\Gamma}(\Theta;k_2,l_2,\sigma),$$

where $\Gamma(\theta;k,l,\sigma)=4\Phi_1^{(R)}(\theta;k,l,\sigma)+4i\Phi_1^{(I)}(\theta;k,l,\sigma)$, $\Phi_1^{(R)}$ *is defined as*

$$\Phi_1^{(R)}(\theta;k,l,\sigma)=\frac{1}{4}\mathbb{E}\Big(\cos(s(\theta+_\pi\Theta_2)+t(\theta-_\pi\Theta_2))+\cos(s(\Theta_2+_\pi\theta)+t(\Theta_2-_\pi\theta))$$

$$-2\cos(s(\theta +_\pi \Theta_2) + t(\Theta_3 -_\pi \Theta_4)) - \cos(s(\Theta_2 +_\pi \Theta_3) + t(\theta -_\pi \Theta_4))$$
$$+2\cos(s(\Theta_2 +_\pi \Theta_3) + t(\Theta_2 -_\pi \Theta_3)) - \cos(s(\Theta_2 +_\pi \Theta_3) + t(\Theta_4 -_\pi \theta))\Big),$$

and $\Phi_1^{(I)}$ *is defined analogously using the sine function. Then it holds that*

$$S_{n,\omega} \xrightarrow[n\to\infty]{\mathcal{D}} \|\mathcal{G}\|^2_{l^2_\omega(\mathbb{Z}^2)}.$$

Proof. Since the test statistic $S_{n,\omega}$ can be represented as $S_{n,\omega} = ||D_n||^2_{l^2_\omega(\mathbb{Z}^2)}$ where

$$D_n(k,l) = \sqrt{n}(\varphi_n(k,l) - \varphi_n(k,0)\varphi_n(0,l)).$$

it is convenient to rewrite D_n in the following form

$$\begin{aligned}
D_n(k,l) =& \frac{\sqrt{n}}{n^4} \sum_{i_1=1}^n \sum_{i_2=1}^n \sum_{i_3=1}^n \sum_{i_4=1}^n \Big(\cos(k(\Theta_{i_1} +_\pi \Theta_{i_2}) + l(\Theta_{i_1} -_\pi \Theta_{i_2})) \\
&+ i\sin(k(\Theta_{i_1} +_\pi \Theta_{i_2}) + l(\Theta_{i_1} -_\pi \Theta_{i_2})) \\
&- (\cos(k(\Theta_{i_1} +_\pi \Theta_{i_2})) + i\sin(k(\Theta_{i_1} +_\pi \Theta_{i_2}))) \\
&\times (\cos(l(\Theta_{i_3} -_\pi \Theta_{i_4})) + i\sin(l(\Theta_{i_3} -_\pi \Theta_{i_4})))\Big) \\
=& \sqrt{n}\Big(\frac{1}{n^4} \sum_{i_1=1}^n \sum_{i_2=1}^n \sum_{i_3=1}^n \sum_{i_4=1}^n (\Phi^{(R)}(\Theta_{i_1}, \Theta_{i_2}, \Theta_{i_3}, \Theta_{i_4}; k, l) \\
&+ i\Phi^{(I)}(\Theta_{i_1}, \Theta_{i_2}, \Theta_{i_3}, \Theta_{i_4}; k, l)\Big)
\end{aligned}$$

where

$$\begin{aligned}
\Phi^{(R)}(\theta_{i_1}, \theta_{i_2}, \theta_{i_3}, \theta_{i_4}; k, l) =& \cos(k(\theta_{i_1} +_\pi \theta_{i_2}) + l(\theta_{i_1} -_\pi \theta_{i_2})) \\
&- \cos(k(\theta_{i_1} +_\pi \theta_{i_2}))\cos(l(\theta_{i_3} -_\pi \theta_{i_4})) \\
&+ \sin(k(\theta_{i_1} +_\pi \theta_{i_2}))\sin(l(\theta_{i_3} -_\pi \theta_{i_4})) \\
=& \cos(k(\theta_{i_1} +_\pi \theta_{i_2}) + l(\theta_{i_1} -_\pi \theta_{i_2})) \\
&- \cos(k(\theta_{i_1} +_\pi \theta_{i_2}) + l(\theta_{i_3} -_\pi \theta_{i_4}))
\end{aligned}$$

and similarly

$$\begin{aligned}
\Phi^{(I)}(\theta_{i_1}, \theta_{i_2}, \theta_{i_3}, \theta_{i_4}; k, l) =& \sin(k(\theta_{i_1} +_\pi \theta_{i_2}) + l(\theta_{i_1} -_\pi \theta_{i_2})) \\
&- \sin(k(\theta_{i_1} +_\pi \theta_{i_2}) + l(\theta_{i_3} -_\pi \theta_{i_4})).
\end{aligned}$$

Since $\Phi^{(R)}$ and $\Phi^{(I)}$ are not symmetric, appropriate symmetrized kernels will be

considered

$$\Phi_s^{(R)}(\theta_{i_1},\theta_{i_2},\theta_{i_3},\theta_{i_4};k,l)=\frac{1}{4!}\sum_{\pi\in\Pi(4)}\Phi^{(R)}(\theta_{\pi(i_1)},\theta_{\pi(i_2)},\theta_{\pi(i_3)},\theta_{\pi(i_4)};k,l)$$

and

$$\Phi_s^{(I)}(\theta_{i_1},\theta_{i_2},\theta_{i_3},\theta_{i_4};k,l)=\frac{1}{4!}\sum_{\pi\in\Pi(4)}\Phi^{(I)}(\theta_{\pi(i_1)},\theta_{\pi(i_2)},\theta_{\pi(i_3)},\theta_{\pi(i_4)};k,l).$$

where $\Pi(4)$ is the set of all permutations of the set $\{i_1,i_2,i_3,i_4\}$.

Using the Hoeffding representations for V-statistics, one obtains

$$D_n(k,l)=\sqrt{n}\left(\frac{4}{n}\sum_{i=1}^{n}\Phi_1^{(R)}(\Theta_{i_1};k,l,\sigma)+r_{1n}(k,l)+i\Big(\frac{4}{n}\sum_{i=1}^{n}\Phi_1^{(I)}(\Theta_{i_1};k,l,\sigma)+r_{2n}(k,l)\Big)\right)$$

where $|r_{1n}(k,l)+ir_{2n}(k,l)|=o_{\mathbb{P}}(\frac{1}{\sqrt{n}})$ and $\Phi_1^{(R)}(\theta;k,l,\sigma)$ and $\Phi_1^{(I)}(\theta;k,l,\sigma)$ are the first projections of the kernels $\Phi_s^{(R)}$ and $\Phi_s^{(I)}$, respectively, equal to

$$\begin{aligned}\Phi_1^{(R)}(\theta;k,l,\sigma)=\frac{1}{4}\mathbb{E}\Big(&\cos(k(\theta+_\pi\Theta_2)+l(\theta-_\pi\Theta_2))-2\cos(k(\theta+_\pi\Theta_2)+l(\Theta_3-_\pi\Theta_4))\\&+\cos(k(\Theta_2+_\pi\theta)+l(\Theta_2-_\pi\theta))+2\cos(k(\Theta_2+_\pi\Theta_3)+l(\Theta_2-_\pi\Theta_3))\\&-\cos(k(\Theta_2+_\pi\Theta_3)+l(\theta-_\pi\Theta_4))-\cos(k(\Theta_2+_\pi\Theta_3)+l(\Theta_4-_\pi\theta))\Big)\end{aligned}$$

and

$$\begin{aligned}\Phi_1^{(I)}(\theta;k,l,\sigma)=\frac{1}{4}\mathbb{E}\Big(&\sin(k(\theta+_\pi\Theta_2)+l(\theta-_\pi\Theta_2))-2\sin(k(\theta+_\pi\Theta_2)+l(\Theta_3-_\pi\Theta_4))\\&+\sin(k(\Theta_2+_\pi\theta)+l(\Theta_2-_\pi\theta))+2\sin(k(\Theta_2+_\pi\Theta_3)+l(\Theta_2-_\pi\Theta_3))\\&-\sin(k(\Theta_2+_\pi\Theta_3)+l(\theta-_\pi\Theta_4))-\sin(k(\Theta_2+_\pi\Theta_3)+l(\Theta_4-_\pi\theta))\Big).\end{aligned}$$

It has been numerically shown that these projections are nonconstant functions. The first projections corresponding to several values of the parameter σ are displayed in Figures 4.1 and 4.2.
Hence,

$$D_n(k,l)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Gamma(\Theta_{i_1};k,l,\sigma)+\sqrt{n}(r_{1n}(k,l)+ir_{2n}(k,l))$$

where $\Gamma(\theta_{i_1};k,l,\sigma)=4\Phi_1^{(R)}(\theta_{i_1};k,l,\sigma)+4i\Phi_1^{(I)}(\theta_{i_1};k,l,\sigma)$. Under the null hypothesis, the expectation of both kernels equals zero, which implies that $\mathbb{E}\Gamma(\Theta_1;k,l,\sigma)=0$. Since the functions $\Phi_1^{(R)}$ and $\Phi_1^{(I)}$ are uniformly bounded,

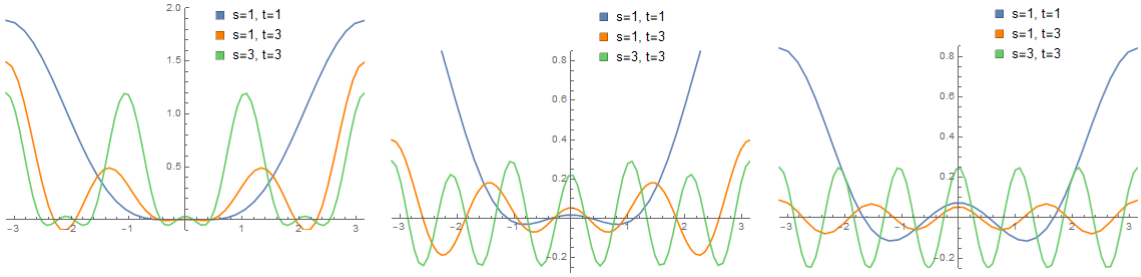


Figure 4.1: The first projection $\Phi_1^{(R)}$ for $\sigma = 0.2, 0.5$ and 0.8 (from left to right)

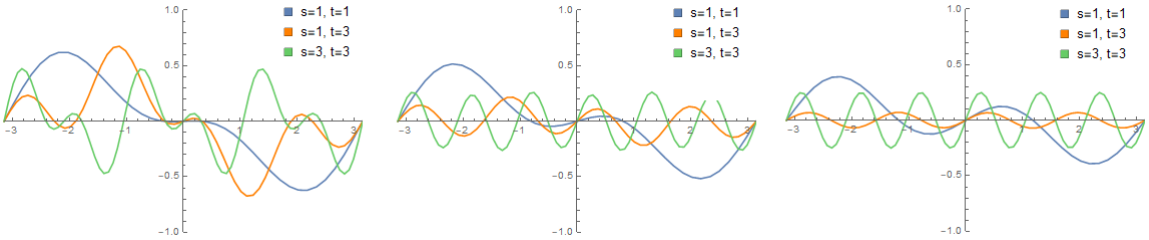


Figure 4.2: The first projection $\Phi_1^{(I)}$ for $\sigma = 0.2, 0.5$ and 0.8 (from left to right)

there exists a constant $C > 0$ such that

$$\mathbb{E}|\Gamma(\Theta_1; k, l, \sigma)|^2 \leq C, \quad (k, l) \in \mathbb{Z}^2.$$

Therefore, using the Tonelli theorem and the condition (4.8), it holds

$$\mathbb{E}\|\Gamma(\Theta_1; \cdot, \cdot, \sigma)\|^2_{l^2_\omega(\mathbb{Z}^2)} = \sum_{k,l=-\infty}^{\infty} \mathbb{E}|\Gamma(\Theta_1; k, l, \sigma)|^2 \omega(k, l) \leq C \sum_{k,l=-\infty}^{\infty} \omega(k, l) < \infty.$$

Now, applying the central limit theorem in Hilbert spaces, it follows that the limiting process is a centered Gaussian process with a specified covariance kernel

$$K_{\mathcal{G}}((k_1, l_1), (k_2, l_2); \sigma) = \mathbb{E}\Gamma(\Theta_1; k_1, l_1, \sigma)\overline{\Gamma}(\Theta_1; k_2, l_2, \sigma). \tag{4.9}$$

The statement of the theorem is an immediate consequence of the continuous mapping theorem. □

Theorem 4.3 (Halaj, 2025). *Assume that the positive weight function ω satisfies*

$$\sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} \omega(k, l) < \infty.$$

Then it holds,

$$\frac{S_{n,\omega}}{n} \xrightarrow[n\to\infty]{a.s.} S := \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} |\varphi(k,l) - \varphi(k,0)\varphi(0,l)|^2 \omega(k,l).$$

Moreover, S is strictly positive whenever $\Theta_1 +_\pi \Theta_2$ and $\Theta_1 -_\pi \Theta_2$ are dependent. Consequently, the test that rejects H_0 for sufficiently large values of $S_{n,\omega}$ is strongly consistent against all fixed alternatives for which this dependence holds.

Proof. According to the strong law of large numbers, for every $(k,l) \in \mathbb{Z}^2$, the joint and marginal empirical characteristic functions converge almost surely to their corresponding theoretical characteristic functions. Using the fact that

$$|\varphi_n(k,l) - \varphi_n(k,0)\varphi_n(0,l)|^2 \leq 4, \quad k,l \in \mathbb{Z}$$

together with the summability of ω over $\mathbb{Z}^2$, the statement of the theorem follows by applying Lebesgue's dominated convergence theorem. □

4.3 Power study

In this section, the performance of the proposed test through an extensive power comparison is examined. Two scenarios are considered: testing the hypothesis for a wrapped normal distribution centered at zero, and a more general case without restrictions. In both settings, the power of $S_{n,\omega}$ is analyzed using the weight function defined below.

Following the approach outlined in García-Portugués et al., 2024, a weight function decomposed as $\omega(k,l) = f(k)f(l)$ is considered. Here, f is a symmetric function about zero. The proposed weight function is particularly convenient, since it leads to a closed-form expression for the test statistic. With this choice, $S_{n,\omega}$ can be rewritten as

$$\begin{aligned}
S_{n,\omega} =& \frac{1}{n^3} \sum_{\substack{i_j=1,\dots,n \\ j\in I_4}} J^{(f)}\left(\Theta_{i_1} +_\pi \Theta_{i_2} - (\Theta_{i_3} +_\pi \Theta_{i_4})\right) J^{(f)}\left(\Theta_{i_1} -_\pi \Theta_{i_2} - (\Theta_{i_3} -_\pi \Theta_{i_4})\right) \\
&+ \frac{1}{n^7} \sum_{\substack{i_j=1,\dots,n \\ j\in I_8}} J^{(f)}\left(\Theta_{i_1} +_\pi \Theta_{i_2} - (\Theta_{i_3} +_\pi \Theta_{i_4})\right) J^{(f)}\left(\Theta_{i_5} -_\pi \Theta_{i_6} - (\Theta_{i_7} -_\pi \Theta_{i_8})\right) \\
&- \frac{2}{n^5} \sum_{\substack{i_j=1,\dots,n \\ j\in I_6}} J^{(f)}\left(\Theta_{i_1} +_\pi \Theta_{i_2} - (\Theta_{i_3} +_\pi \Theta_{i_4})\right) J^{(f)}\left(\Theta_{i_1} -_\pi \Theta_{i_2} - (\Theta_{i_5} -_\pi \Theta_{i_6})\right),
\end{aligned}$$

where $I_k = \{1, \cdots, k\}$ and

$$J^{(f)}(x) := \sum_{k=-\infty}^{\infty} \cos(kx) f(k). \tag{4.10}$$

v (4.10) is the real part of its characteristic function evaluated at x and is equal to

$$J^{(f)}(x) = \cos(\lambda \sin x) e^{\lambda(\cos x - 1)}. \tag{4.11}$$

Under this setting, the analysis is performed for $\lambda \in \{0.1, 0.5, 1, 2\}$. In the sequel, the corresponding test statistic will be denoted by $S_{n,\lambda}$.
For comparison purposes, the following competing tests are included in the study:

- Kuiper test with the test statistic

  $$K_n = \max_{1 \le j \le n} \left\{ U_{(j)} - \frac{j-1}{n} \right\} + \max_{1 \le j \le n} \left\{ \frac{j}{n} - U_{(j)} \right\}$$

  where $U_{(j)}$ are order statistics of $U_j = F(\theta_j; \hat{\mu}_n, \hat{\sigma}_n)$ and $F(\cdot; \hat{\mu}_n, \hat{\sigma}_n)$ is the distribution function of $\mathcal{WN}(\hat{\mu}_n, \hat{\sigma}_n)$;

- Watson test with the test statistic

  $$W_n = \frac{1}{12n} + \sum_{j=1}^{n} \left(\left(U_{(j)} - \frac{2j-1}{2n} \right) - \left(\overline{U} - \frac{1}{2} \right) \right)^2,$$

  where $U_{(j)}$ are order statistics of $U_j = F(\theta_j; \hat{\mu}_n, \hat{\sigma}_n)$, $\overline{U} = \frac{1}{n} \sum_{j=1}^{n} U_j$ and $F(\cdot; \hat{\mu}_n, \hat{\sigma}_n)$ is defined as above;

- Test statistic proposed in Jammalamadaka et al., 2019

  $$C_{n,\lambda} = n \sum_{r=0}^{\infty} \left(\varphi_n(r) - \varphi(r; \hat{\beta}_n) \right)^2 p(r; \lambda).$$

  Here, $p(\cdot;\ \lambda)$ represents pmf of the Poisson distribution with mean λ. In Jammalamadaka et al., 2019, $\lambda = 0.5$ was identified as the optimal value for testing gof to the von Mises distribution. Although similar behaviour may be expected when testing gof to the wrapped normal distribution, several values of λ are considered to determine whether the same choice remains appropriate in this context. The terms $\varphi_n(\cdot)$ and $\varphi(\cdot\ ; \hat{\beta}_n)$ denote the empirical distribution function based on the observed sample and theoretical

characteristic functions of the wrapped normal distribution, respectively, where $\hat{\beta}_n$ represents the mle of the parameter vector $\beta = (\mu, \sigma)$.

Since test statistics are not distribution-free, both empirical sizes and powers are approximated using the warp-speed bootstrap algorithm presented in Giacomini et al., 2013 with $N = 5000$ replications. The simulation study is conducted for sample sizes $n = 20$ and $n = 50$, under the following alternative models:

- a circular uniform distribution $\mathcal{U}[-\pi, \pi)$

$$f(\theta) = \frac{1}{2\pi}, \ -\pi \leq \theta < \pi;$$

- a wrapped Cauchy distribution $\mathcal{WC}(\mu, \rho)$

$$f(\theta) = \frac{1}{2\pi} \frac{1 - \rho^2}{1 + \rho^2 - 2\rho\cos(\theta - \mu)}, \ -\pi \leq \theta, \mu < \pi, \ 0 < \rho < 1;$$

- a von Mises distribution $\mathcal{VM}(\mu, k)$

$$f(\theta) = \frac{1}{2\pi I_0(k)} e^{k\cos(\theta - \mu)}, \ -\pi \leq \theta, \mu < \pi, \ k > 0;$$

where

$$I_0(k) = \sum_{i=0}^{\infty} \frac{1}{(i!)^2} \left(\frac{k}{2}\right)^{2i}$$

is the modified Bessel function of the first kind;

- a Cartwright's power-of-cosine distribution $\mathcal{CW}(\mu, \zeta)$

$$f(\theta) = \frac{2^{\frac{1}{\zeta} - 1} \Gamma^2\left(\frac{1}{\zeta} + 1\right) (1 + \cos(\theta - \mu))^{\frac{1}{\zeta}}}{\pi \Gamma\left(\frac{2}{\zeta} + 1\right)}, \ -\pi \leq \theta, \mu < \pi, \ \zeta > 0,$$

where $\Gamma(\cdot)$ is the gamma function;

- a Jones-Pewsey distribution $\mathcal{JP}(\mu, \kappa, \psi)$

$$f(\theta) = \frac{(\cosh(\kappa\psi) + \sinh(\kappa\psi)\cos(\theta - \mu))^{\frac{1}{\psi}}}{2\pi P_{\frac{1}{\psi}}(\cosh(\kappa\psi))}, \ -\pi \leq \theta, \mu < \pi, \ \kappa \geq 0, \ \psi \in \mathbb{R},$$

where

$$P_{\frac{1}{\psi}}(x) = \frac{1}{\pi} \int_0^{\pi} \left(x + \sqrt{x^2 - 1}\cos t\right)^{1/\psi} dt, \quad x \geq 1,$$

is the associated Legendre function of the first kind with degree $\frac{1}{\psi}$ and order 0;

- a Batschelet distribution $\mathcal{B}a(\mu, \kappa, \nu)$

$$f(\theta) = \frac{1}{B_0(\kappa, \nu)} e^{\kappa \cos((\theta-\mu)+\nu \sin(\theta-\mu))}, \ -\pi \leq \theta, \mu < \pi, \ \kappa \geq 0, \ -1 \leq \nu \leq 1,$$

where

$$B_p(\kappa, \nu) = \int_0^{2\pi} \cos(p\theta) e^{\kappa \cos(\theta+\nu \sin\theta)} d\theta, \ p = 0, 1, ...$$

- a wrapped α-stable distribution $\mathcal{WS}(\mu, \tau, \alpha, \beta)$

$$f(\theta) = \frac{1}{2\pi} + \sum_{k=1}^{\infty} \frac{e^{-\tau^\alpha k^\alpha}}{\pi} \cos\Big(k(\theta-\mu) - \tau^\alpha k^\alpha \beta \tan\frac{\pi\alpha}{2}\Big), \ -\pi \leq \theta, \mu < \pi,$$

with $\alpha \in (0, 1) \cup (1, 2]$, $\tau \geq 0$ and $|\beta| \leq 1$;

- a mixture of two wrapped normal distributions

$$p\mathcal{WN}(\mu_1, \sigma_1) + (1-p)\mathcal{WN}(\mu_2, \sigma_2), \quad p \in (0, 1);$$

- a mixture of two von Mises distributions

$$p\mathcal{VM}(\mu_1, \kappa_1) + (1-p)\mathcal{VM}(\mu_2, \kappa_2), \quad p \in (0, 1).$$

In the simulations, the following notation is used.

Notation	Mixture
$\mathcal{WN}_{M1}(p)$	$p\mathcal{WN}(0, 0.2) + (1-p)\mathcal{WN}(0, 0.8)$
$\mathcal{WN}_{M2}(p)$	$p\mathcal{WN}(0, 0.2) + (1-p)\mathcal{WN}(0, 2)$
$\mathcal{WN}_{M3}(p)$	$p\mathcal{WN}(-\pi, 0.2) + (1-p)\mathcal{WN}(\frac{\pi}{2}, 0.8)$
$\mathcal{VM}_{M1}(p)$	$p\mathcal{VM}(0, 0.5) + (1-p)\mathcal{VM}(0, 5)$
$\mathcal{VM}_{M2}(p)$	$p\mathcal{VM}(0, 1) + (1-p)\mathcal{VM}(0, 5)$
$\mathcal{VM}_{M3}(p)$	$p\mathcal{VM}(-\pi, 5) + (1-p)\mathcal{VM}(\frac{\pi}{2}, 5)$
$\mathcal{VM}_{M4}(p)$	$p\mathcal{VM}(-\pi, 3) + (1-p)\mathcal{VM}(0.62\pi, 3)$
$\mathcal{VM}_{M5}(p)$	$p\mathcal{VM}(-\pi, 8) + (1-p)\mathcal{VM}(-\pi, 0.1)$

For all other distributions that are not mixtures, alternatives with the location parameter fixed at 0 were employed in Tables 4.1 and 4.4. Accordingly, the location parameter μ has been omitted from the notation.

4.3.1 The first case

This subsection focuses on testing the null hypothesis for the wrapped normal distribution, where the distribution is assumed to be centered at zero, i.e.:

$$H_0 : \Theta \sim \mathcal{WN}(0, \sigma), \quad \text{for some } \sigma > 0,$$

against the alternative:

$$H_1 : \Theta \not\sim \mathcal{WN}(0, \sigma) \quad \text{for all } \sigma > 0.$$

To evaluate the finite-sample behavior of the test in this setting, empirical power values are obtained using the warp-speed bootstrap procedure described below.

Algorithm 3 Warp-speed bootstrap algorithm

1: Generate a random sample $\Theta_1, \Theta_2, \ldots, \Theta_n$ from the alternative distribution F;
2: Calculate the estimate $\hat{\sigma}_n$ of unknown parameter σ and value of the test statistic $T_n = T_n(\theta_1, \theta_2, ...\theta_n; \hat{\sigma}_n)$ based on the observed sample;
3: Generate a bootstrap sample $\Theta_1^*, \Theta_2^*, ..., \Theta_n^*$ where each $\Theta_i^* \sim \mathcal{WN}(0, \hat{\sigma}_n)$;
4: Using the observed sample $(\theta_1^*, \ldots, \theta_n^*)$, calculate the bootstrap estimate $\hat{\sigma}_n^*$ of the parameter σ;
5: Compute the value of the bootstrap test statistic $T_n^* = T_n(\theta_1^*, \theta_2^*, ..., \theta_n^*; \hat{\sigma}_n^*)$;
6: Repeat the previous steps N times to generate two separate sequences of test statistics: $\{T_n^{(1)}, T_n^{(2)}, ..., T_n^{(N)}\}$ and $\{T_n^{*(1)}, T_n^{*(2)}, ..., T_n^{*(N)}\}$;
7: Reject the null hypothesis for $j = 1, ..., N$ if $T_n^{(j)} > c_\alpha$, where c_α denotes the $(1 - \alpha)100\%$ quantile of the empirical distribution of the bootstrap test statistics $(T_n^{*(j)}, j = 1, ..., N)$.

The estimation of σ will be based on the maximum likelihood estimator; further details are provided in Subsection 4.3.2. The percentages of empirical sizes (shown in the first four rows) and empirical powers are presented in Table 4.1. The highest powers are highlighted in bold for emphasis. The calculations are based on $N = 5000$ replications. As a result, the standard error for the empirical powers is not greater than $0.5/\sqrt{5000}$. According to Halaj et al., 2024, gray-shaded cells represent power percentages within $1.96 \times 0.5/\sqrt{5000} \cdot 100\% = 1\%$ of the maximum value, including the maximum itself.

Regarding the significance level, all tests are well-calibrated. The results further reveal that no single testing procedure consistently attains the highest power across all considered alternatives, which is in accordance with the findings re-

	$n = 20$										$n = 50$									
Dist.	$S_{n,0.1}$	$S_{n,0.5}$	$S_{n,1}$	$S_{n,2}$	K_n	W_n	$C_{n,0.1}$	$C_{n,0.5}$	$C_{n,1}$	$C_{n,2}$	$S_{n,0.1}$	$S_{n,0.5}$	$S_{n,1}$	$S_{n,2}$	K_n	W_n	$C_{n,0.1}$	$C_{n,0.5}$	$C_{n,1}$	$C_{n,2}$
$\mathcal{WN}(0.2)$	4	4	4	4	5	6	5	5	5	5	4	4	4	4	5	5	5	5	5	5
$\mathcal{WN}(0.5)$	4	4	5	5	5	6	5	5	5	6	4	5	5	6	5	5	5	6	6	6
$\mathcal{WN}(0.8)$	4	5	5	5	5	5	5	5	6	6	4	5	5	5	5	5	6	6	6	5
$\mathcal{WN}(2)$	5	5	5	5	4	5	4	5	5	5	5	5	5	5	4	5	5	5	5	5
$\mathcal{U}[-\pi,\pi)$	5	5	5	5	**6**	**6**	**6**	**6**	**6**	5	**6**	**6**	**6**	5	**6**	**6**	**6**	**6**	**6**	**6**
$\mathcal{VM}(0.5)$	**5**	**5**	**5**	**5**	4	4	4	4	4	4	5	**6**	**6**	5	4	5	4	4	5	5
$\mathcal{VM}(1)$	**6**	**6**	**6**	4	4	5	5	5	5	**6**	**9**	**9**	8	6	6	6	4	5	6	6
$\mathcal{VM}(2)$	**7**	**7**	6	5	**7**	6	4	5	6	**7**	**12**	11	9	6	10	9	5	6	8	8
$\mathcal{VM}(5)$	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**7**	5	4	5	5	5	5	5	5	5
$\mathcal{WC}(0.2)$	**5**	**5**	**5**	**5**	4	4	4	4	4	**5**	**6**	5	**6**	5	5	4	4	4	4	5
$\mathcal{WC}(0.5)$	**16**	**16**	15	12	10	9	4	4	8	12	**43**	**43**	41	28	27	27	4	11	26	36
$\mathcal{WC}(0.8)$	49	52	52	42	64	**65**	2	16	37	55	83	88	91	88	97	**98**	5	55	80	94
$\mathcal{CW}(1)$	**7**	**7**	**7**	6	6	6	6	6	6	6	**9**	**9**	8	6	7	7	6	7	7	7
$\mathcal{CW}(2)$	**8**	**8**	**8**	6	7	5	7	7	7	7	**12**	**12**	**12**	7	8	6	6	7	8	9
$\mathcal{CW}(5)$	**6**	**6**	**6**	5	5	5	5	5	**6**	5	9	**10**	9	7	7	6	6	6	7	8
$\mathcal{JP}$(2,-1)	46	48	49	38	58	**59**	2	13	34	53	85	88	90	83	96	**98**	6	58	82	94
$\mathcal{JP}$(2,1)	**7**	**7**	6	5	6	6	6	6	**7**	6	**8**	**8**	**8**	5	7	6	7	7	7	7
$\mathcal{JP}$(2,2)	**8**	**8**	7	6	6	5	7	7	7	7	**12**	**12**	**12**	7	8	6	6	7	8	10
$\mathcal{B}a$(3,-1)	**16**	**16**	**16**	8	11	10	8	9	11	13	**34**	33	30	16	19	19	8	12	19	25
$\mathcal{B}a$(3,0.5)	**16**	15	11	6	12	12	4	6	9	11	**32**	25	14	7	21	22	4	8	13	20
$\mathcal{B}a$(3,1)	19	**20**	18	6	19	19	3	10	13	17	**39**	37	23	9	**39**	**39**	5	17	26	33
$\mathcal{B}a$(1,-1)	**10**	**10**	9	6	7	7	7	8	9	8	**18**	**18**	**18**	11	11	10	7	9	13	14
$\mathcal{B}a$(1,0.5)	**14**	**14**	13	8	9	8	3	4	7	11	**31**	**31**	30	20	22	20	4	10	19	26
$\mathcal{B}a$(1,1)	**29**	27	27	16	17	16	3	6	14	22	**66**	**66**	**66**	49	48	48	4	24	49	59
$\mathcal{WS}$(1,0.5,0.3)	22	22	**24**	21	20	17	6	8	16	**24**	51	53	58	56	57	55	21	37	52	**63**
$\mathcal{WS}$(1,0.5,0.5)	22	23	25	23	29	28	16	21	26	**31**	51	54	58	57	74	74	48	62	72	**76**
$\mathcal{WS}$(1,0.5,0.8)	21	22	25	24	**46**	**46**	36	41	43	41	50	53	58	59	**92**	91	81	88	90	89
$\mathcal{VM}_{M1}(0.2)$	**32**	29	25	10	26	25	2	7	18	26	**64**	62	52	23	60	62	4	28	48	61
$\mathcal{VM}_{M2}(0.2)$	**24**	21	15	7	17	18	3	6	12	17	**45**	40	31	12	39	40	4	13	27	37
$\mathcal{WN}_{M1}(0.5)$	18	23	28	**29**	**29**	**29**	4	4	7	18	43	57	67	70	74	**76**	4	6	19	55
$\mathcal{WN}_{M1}(0.8)$	31	36	**39**	27	**39**	**39**	3	5	9	16	59	65	66	59	79	**82**	3	8	23	53
$\mathcal{WN}_{M2}(0.2)$	15	16	**19**	18	11	10	4	5	9	16	33	35	40	**45**	33	25	4	10	24	43
$\mathcal{WN}_{M2}(0.8)$	65	71	70	63	**88**	**88**	2	39	63	80	94	96	96	95	**100**	**100**	16	83	96	99

Table 4.1: Empirical sizes and powers (expressed as percentages) for different sample sizes at significance level 0.05 (the first case: $\mu = 0$ known).

ported in Janssen, 2000. For the newly proposed test, the power remains comparable across different values of λ, except at $\lambda = 2$, where a substantial decline in power is observed. This tendency is consistent with the results of García-Portugués et al., 2024. A natural explanation is that, when $\lambda = 2$, the function (4.11) is not everywhere non-negative. Consequently, positive and negative pairwise contributions may partially offset each other, leading to a loss of sensitivity and, hence, lower power. On the other hand, tests K_n and W_n have similar power and typically outperform $C_{n,\lambda}$ for almost all alternatives. Overall, the new test performs competitively compared to the other proposed tests.

Although the circular uniform alternative is formally excluded by the characterization assumptions, it was included in the simulation study to examine the behavior of the tests in this boundary case. Under circular uniformity, the circular sum and circular difference are also independent, and hence the proposed test statistics do not exhibit a systematic departure from their null behavior. Consequently, the empirical powers are close to the nominal significance level. The same pattern is observed for the competing tests. From another perspective, this behavior is natural, since the circular uniform distribution represents a degenerate limiting case of the wrapped normal distribution as $\sigma \to \infty$.

On the other hand, since the von Mises and von Mises-related alternatives are close to the wrapped normal distribution, deviations from wrapped normality are more difficult to detect, leading to generally lower empirical powers. This suggests that these distributions are more challenging to distinguish from the null hypothesis in such scenarios. To further explore this issue, power results for larger sample sizes are included for two illustrative parameter settings, showing which sample size is sufficient to achieve reasonable statistical power. The results presented in Table 4.2 indicate that a sample size of approximately $n = 100$ is generally sufficient to achieve adequate power.

Alt.	n	$S_{n,0.1}$	$S_{n,0.5}$	$S_{n,1}$	$S_{n,2}$	K_n	W_n	$C_{n,0.1}$	$C_{n,0.5}$	$C_{n,1}$	$C_{n,2}$
$\mathcal{VM}(1)$	70	10	10	9	7	7	7	5	5	7	8
	80	11	11	10	8	7	8	4	5	7	8
	100	14	14	12	9	10	9	5	7	10	11
	150	17	17	17	11	11	11	5	8	11	13
$\mathcal{VM}(2)$	70	16	13	10	6	11	11	5	7	9	11
	80	17	14	11	6	13	13	5	7	10	12
	100	23	18	14	8	14	14	5	8	12	15
	150	30	25	17	10	19	20	5	10	17	20

Table 4.2: Empirical powers (expressed as a percentage) for larger sample sizes against von Mises alternative at $\alpha = 0.05$ (The first case: $\mu = 0$ (known))

4.3.2 The second case

In this subsection, the general case of testing the null hypothesis:

$$H_0 : \Theta \sim \mathcal{WN}(\mu, \sigma), \text{ for some } -\pi \leq \mu < \pi, \ \sigma > 0,$$

against the alternative hypothesis:

$$H_1 : \Theta \not\sim \mathcal{WN}(\mu, \sigma), \text{ for all } -\pi \leq \mu < \pi, \ \sigma > 0$$

is considered. The warp-speed bootstrap procedure closely follows the approach proposed in the first case. In steps 2 and 4, an additional estimate $\hat{\mu}_n$ and a bootstrap estimate $\hat{\mu}_n^*$ of the parameter μ are calculated, respectively. In step 3, a bootstrap sample is generated from the $\mathcal{WN}(\hat{\mu}_n, \hat{\sigma}_n)$ distribution. Parameter estimation is performed using the algorithm implemented in the function `mle.wrappednormal` from the R package `circular` (Agostinelli and Lund, 2022). This algorithm employs an iterative reweighted maximum likelihood (IRML) procedure (Agostinelli, 2007) to compute the mle. The initial values are derived from estimates obtained via the method of moments:

$$\mu_0 = \text{atan2}\left(\sum_{i=1}^{n} \sin(\theta_i), \sum_{i=1}^{n} \cos(\theta_i)\right) \quad \text{and} \quad \sigma_0 = \sqrt{-2 \log R},$$

where R is the resultant length

$$R = \sqrt{\left(\frac{1}{n}\sum_{i=1}^{n} \cos(\theta_i)\right)^2 + \left(\frac{1}{n}\sum_{i=1}^{n} \sin(\theta_i)\right)^2}.$$

It is worth noting that the mle $\hat{\mu}_n$ of the parameter μ exhibits equivariance, whereas the mle $\hat{\sigma}_n$ of the parameter σ remains invariant under operations (4.2) and (4.3). Consequently, when the data are centered as $\theta'_j = \theta_j -_\pi \hat{\mu}_n$ for $j = 1, \ldots, n$, the distribution of any test statistic that depends on $\hat{\mu}_n$ through θ'_j will be unaffected by the particular value of μ. Thus, without loss of generality, it is assumed that $\mu = 0$.

The results reported in Table 4.4 indicate that the tests remain well calibrated in this setting. Their overall performance is similar to that observed in the first case. Among the different choices of the parameter λ for test statistic $S_{n,\lambda}$, the case $\lambda = 2$ again yields the lowest power, which further supports its less favorable behavior. By contrast, the $C_{n,\lambda}$ test demonstrates improved performance, at-

taining powers comparable to those of the newly proposed test in most cases and outperforming it under some alternatives. Moreover, consistent with the findings reported in Jammalamadaka et al., 2019 concerning gof testing for the von Mises distribution, the results obtained here indicate that $\lambda = 0.5$ provides the best overall performance in the present setting. Nevertheless, the proposed test performs consistently well across a broad range of alternatives and remains competitive. Under the circular uniform alternative, it exhibits the same behavior as in the previous case. In the same manner, Table 4.3 reports the powers of the tests for larger sample sizes under von Mises alternatives. Although departures from the null hypothesis can already be detected at smaller sample sizes, the choice $n = 100$ yields more stable and reliable results over a wider range of parameter values.

Alt.	n	$S_{n,0.1}$	$S_{n,0.5}$	$S_{n,1}$	$S_{n,2}$	K_n	W_n	$C_{n,0.1}$	$C_{n,0.5}$	$C_{n,1}$	$C_{n,2}$
$\mathcal{VM}(1)$	70	11	11	11	7	9	10	11	11	11	10
	80	11	11	11	8	10	11	12	12	11	10
	100	13	12	12	9	11	11	13	13	12	11
	150	18	18	18	13	16	17	18	19	19	16
$\mathcal{VM}(2)$	70	16	14	11	7	14	15	20	20	19	15
	80	18	15	12	7	16	17	23	23	21	17
	100	20	18	13	9	18	20	25	26	23	18
	150	32	28	20	10	27	28	35	38	36	28

Table 4.3: Empirical powers (expressed as a percentage) for larger sample sizes against von Mises alternative at $\alpha = 0.05$ (The second case: unknown μ)

An important issue in the practical implementation of the proposed $S_{n,\lambda}$ test is the choice of the tuning parameter λ. A reasonable default is $\lambda = 1$, since the corresponding test statistic exhibits stable and balanced performance across all scenarios considered. Alternatively, λ may be selected by means of a data-driven procedure (see, e.g., Allison and Santana 2015; Tenreiro 2019).

	$n = 20$										$n = 50$									
Dist.	$S_{n,0.1}$	$S_{n,0.5}$	$S_{n,1}$	$S_{n,2}$	K_n	W_n	$C_{n,0.1}$	$C_{n,0.5}$	$C_{n,1}$	$C_{n,2}$	$S_{n,0.1}$	$S_{n,0.5}$	$S_{n,1}$	$S_{n,2}$	K_n	W_n	$C_{n,0.1}$	$C_{n,0.5}$	$C_{n,1}$	$C_{n,2}$
$\mathcal{WN}(0.2)$	4	4	4	5	5	5	4	4	4	4	5	5	5	5	5	5	5	5	5	5
$\mathcal{WN}(0.5)$	4	5	5	**6**	5	5	4	4	5	5	5	5	5	5	5	5	5	5	5	6
$\mathcal{WN}(0.8)$	5	5	**6**	5	5	5	5	5	5	6	5	5	5	5	5	5	5	5	5	5
$\mathcal{WN}(2)$	6	6	5	4	4	3	4	5	5	5	5	5	5	4	4	4	4	5	5	5
$\mathcal{U}[-\pi,\pi)$	**6**	**6**	5	5	3	3	4	5	5	5	**6**	**6**	**6**	5	4	4	5	5	5	5
$\mathcal{VM}(0.5)$	**6**	**6**	5	4	4	4	5	5	5	5	**6**	**6**	**6**	5	5	5	5	**6**	**6**	5
$\mathcal{VM}(1)$	7	7	6	5	6	6	**8**	**8**	**8**	7	**9**	8	8	6	7	8	**9**	**9**	**9**	7
$\mathcal{VM}(2)$	8	7	6	5	7	8	**12**	11	10	9	13	11	9	6	11	12	16	**17**	15	11
$\mathcal{VM}(5)$	**7**	**7**	6	6	5	6	**7**	**7**	**7**	**7**	7	6	5	5	6	6	**9**	8	8	7
$\mathcal{WC}(0.2)$	**6**	**6**	5	4	4	4	5	5	5	5	**6**	**6**	**6**	5	5	5	5	5	**6**	5
$\mathcal{WC}(0.5)$	17	17	16	11	17	17	19	**20**	19	17	42	42	41	27	40	43	41	**44**	**44**	42
$\mathcal{WC}(0.8)$	50	53	52	42	73	**76**	50	58	63	69	85	88	90	87	**99**	**99**	83	91	94	97
$\mathcal{CW}(1)$	**8**	**8**	7	5	6	6	**8**	**8**	7	7	**11**	**11**	10	6	8	8	9	10	10	8
$\mathcal{CW}(2)$	**8**	**8**	**8**	6	5	6	**8**	**8**	**8**	7	**11**	**11**	10	7	7	8	10	10	10	9
$\mathcal{CW}(5)$	**7**	**7**	6	4	4	3	6	6	6	6	**8**	**8**	7	6	5	5	6	7	7	7
$\mathcal{JP}(2,-1)$	49	51	51	38	66	**71**	51	58	63	67	86	89	90	84	98	**99**	84	92	95	97
$\mathcal{JP}(2,1)$	**7**	**7**	**7**	5	6	6	**7**	**7**	**7**	6	**10**	9	9	5	7	7	8	9	8	8
$\mathcal{JP}(2,2)$	**9**	8	8	6	5	6	8	8	8	7	**11**	**11**	10	7	8	8	10	10	9	9
$\mathcal{B}a(3,-1)$	**18**	**18**	17	8	11	13	11	15	16	15	**36**	35	32	15	24	26	20	31	32	29
$\mathcal{B}a(3,0.5)$	17	17	12	6	14	16	18	18	**19**	**19**	33	26	15	7	27	30	35	**36**	**36**	33
$\mathcal{B}a(3,1)$	21	20	19	8	22	**23**	21	22	**23**	**23**	40	38	24	9	43	**45**	41	43	44	**45**
$\mathcal{B}a(1,-1)$	**12**	11	10	6	7	6	10	10	10	9	**17**	**17**	**17**	11	12	13	15	16	16	13
$\mathcal{B}a(1,0.5)$	16	16	15	9	14	15	**19**	18	17	16	34	33	31	19	29	32	33	**35**	33	30
$\mathcal{B}a(1,1)$	29	28	27	17	28	30	31	**33**	32	30	68	69	68	49	64	69	63	**70**	**70**	67
$\mathcal{WS}(1,0.5,0.3)$	25	26	27	22	26	25	22	26	27	**28**	54	56	60	56	60	**61**	46	55	58	**61**
$\mathcal{WS}(1,0.5,0.5)$	25	25	27	21	26	25	23	26	27	**28**	53	54	57	54	58	59	44	54	56	**60**
$\mathcal{WS}(1,0.5,0.8)$	25	25	**27**	21	25	26	22	25	**27**	**27**	54	56	60	56	60	60	47	56	59	**62**
$\mathcal{VM}_{M3}(0.8)$	35	**38**	34	20	26	33	31	36	**38**	37	81	**83**	79	57	70	80	75	**83**	**83**	82
$\mathcal{VM}_{M3}(0.5)$	24	33	**36**	24	26	29	9	21	28	33	64	74	**76**	64	72	**76**	30	62	71	75
$\mathcal{VM}_{M4}(2/3)$	7	**8**	**8**	6	6	7	7	7	7	**8**	12	**13**	**13**	9	10	11	10	12	**13**	**13**
$\mathcal{VM}_{M5}(1/3)$	26	27	**28**	19	23	26	22	27	**28**	**28**	58	59	60	52	61	62	48	58	62	**63**
$\mathcal{WN}_{M3}(0.3)$	27	34	**39**	34	32	30	15	25	32	38	59	70	**79**	**79**	69	70	38	58	67	78
$\mathcal{WN}_{M3}(0.5)$	64	77	**84**	79	78	80	52	68	76	82	98	**100**	**100**	**100**	**100**	**100**	94	99	99	**100**
$\mathcal{WN}_{M3}(0.9)$	60	64	62	44	77	**78**	59	65	69	72	90	90	88	76	**98**	**98**	90	93	94	96

Table 4.4: Empirical sizes and powers (expressed as percentages) for different sample sizes at significance level 0.05 (the second case: unknown μ).

Chapter 5

A new approach to goodness-of-fit testing based on linear combinations of degenerate V-statistics

In nonparametric statistics, no test achieves uniformly superior performance across all possible alternatives. The sensitivity of each test depends on the specific form of the alternative, so a method that is powerful in one context may be less effective in another (see, e.g., Janssen, 2000). Although many nonparametric tests are omnibus, ensuring consistency against any deviation from the null hypothesis, their power in finite samples can vary substantially, which may affect the reliability of empirical results.

To overcome this limitation, researchers often conduct multiple tests on the same dataset to assess the same null hypothesis. One common strategy is to combine the resulting p-values of individual tests, as originally proposed in Fisher's method (Fisher, 1934). Numerous extensions and modifications of this approach have been developed (see, e.g., Van Zwet and Oosterhoff, 1967; Loughin, 2004). Although these techniques are effective when tests are independent, their accuracy tends to decline when tests are correlated. Several approaches to this problem have been proposed in the literature, including those in Yu et al., 2009 and Sexton et al., 2012.

Many test statistics represent specific instances within a broader family of statistics characterized by certain tuning parameters. In these contexts, data-driven methods are commonly employed to select parameters that optimize test power (Allison and Santana, 2015; Tenreiro, 2019; Fernández-de-Marcos and García-Portugués, 2023). However, such approaches are often computationally intensive and impractical for real-world applications. In addition, such procedures are usually developed in settings where the null distribution of the underlying test statistic is parameter-free or where the required critical values can be

precomputed for each tuning parameter. When the null distribution depends on unknown parameters, these critical values must instead be approximated for the estimated parameter values, which may considerably increase the computational burden in practical applications.

An alternative perspective focuses on combining the test statistics themselves. For example, Doksum and Thompson, 1971 introduced optimal linear combinations of sign and Wilcoxon tests to enhance sensitivity for location and symmetry testing. The efficiency and performance of such combined symmetry tests were later investigated in Burgio and Nikitin, 2001 and Burgio and Nikitin, 2003. Similarly, Deshpande and Kochar, 1983 analyzed linear combinations of exponentiality tests designed to detect "new better than used" alternatives. These types of test statistics are related to nondegenerate U- or V-statistics, which are known to be asymptotically normal under the null hypothesis. In contrast, many gof tests, especially those based on L^2-type distance measures, correspond to degenerate V-statistics, sometimes involving estimated parameters.

In Section 1.2.2, a general linear combination of weakly degenerate V-statistics was considered. The limiting null distribution was established under different parameter settings. These include the case without parameter estimation, the case in which the components depend on estimated parameters, and the mixed case where only some components involve estimated parameters. The present chapter focuses on the choice of weights, with particular emphasis on the case in which each component is weighted by the inverse of its corresponding null standard deviation. This weighting scheme ensures that all components contribute on a comparable scale, thereby avoiding dominance by any individual statistic due solely to differences in variability or measurement units.

The proposed approach is designed for implementation in two main settings:

- *Tuning parameter adjustment*: Certain test statistics depend on a tuning parameter (e.g., in weighted L^2-distance tests), which determines the sensitivity of the test to different types of alternatives. Although simulation studies often suggest a recommended tuning parameter, such a choice may favor particular alternatives while reducing power against others. The proposed combination procedure reduces the influence of a single parameter choice by incorporating test statistics corresponding to several parameter values.

- *Complementary power properties*: In some cases, two test statistics exhibit complementary power behavior, where one statistic has lower power against certain alternatives, while another performs better against the same alternatives. Combining such statistics can offset individual weaknesses

and lead to a composite test with more stable power across a wider range of alternatives.

5.1 Weight selection

Let $X_1, \ldots, X_n$ be an i.i.d. sample with distribution F and support $S \subseteq \mathbb{R}^d$. Let

$$\mathcal{F} = \{F(\cdot;\lambda),\ \lambda \in \Lambda\}$$

be a parametric family of distribution functions, where λ is an unknown parameter and $\Lambda \subseteq \mathbb{R}^p$ is the corresponding parameter space. The problem of interest is to test the null hypothesis

$$H_0 : F \in \mathcal{F}$$

against the alternative

$$H_1 : F \notin \mathcal{F}.$$

Consider a collection of gof test statistics $T_n^1, \ldots, T_n^k$, each of which can be represented as a weakly degenerate V-statistic. The proposed procedure constructs a combined statistic as a weighted linear combination of these components, with weights inversely proportional to their limiting or exact standard deviations under the null hypothesis. The precise form of the weights depends on whether the null distribution of the corresponding component statistic is invariant with respect to the parameter λ.

Therefore, the construction is considered separately in two cases:

(i) If the null distribution of T_n^i, $i = 1, \ldots, k$, is free of λ, the combined statistic is constructed as

$$L_n = \sum_{i=1}^{k} \frac{T_n^i}{\sigma_0^i}, \qquad \sigma_0^i = \sqrt{\mathrm{Var}_{H_0}(nT_n^i)}.$$

In practice, the quantities $\mathrm{Var}_{H_0}(nT_n^i)$ are usually not available in closed form and therefore have to be approximated. One possibility is to estimate them by a Monte Carlo procedure for each fixed sample size n. Another possibility is to use the variance of the limiting statistic, either computed from (1.14) or approximated by Monte Carlo simulation. Since the eigenvalues of the corresponding integral operator often decrease rapidly, the series in (1.14) can usually be truncated after a relatively small number of dominant terms.

Moreover, in many cases the largest eigenvalue accounts for a substantial part of the total variance. This motivates the alternative combined statistic

$$\tilde{L}_n = \sum_{i=1}^{k} \frac{T_n^i}{\nu_1^i},$$

where ν_1^i denotes the largest eigenvalue of the integral operator associated with the limiting distribution of nT_n^i. Although the exact computation of this eigenvalue may be difficult, it can be efficiently approximated (see, e.g., Ebner et al. 2025 for further details).

(ii) The parameter-dependent case arises when the null distribution of at least one statistic T_n^i, $i = 1, \ldots, k$, depends on the unknown parameter λ. Thus, the corresponding standard deviation has to be treated as a parameter-dependent quantity, denoted by $\sigma_0^i(\lambda)$. If an explicit expression for $\sigma_0^i(\lambda)$ is available, the natural choice is to use the weight $1/\sigma_0^i(\hat{\lambda}_n)$, where $\hat{\lambda}_n$ is a consistent estimator of λ. When such a closed-form expression is not available, $\sigma_0^i(\hat{\lambda}_n)$ may be approximated by bootstrap methods. Since, in this setting, the null distribution of the combined statistic L_n also depends on λ, resampling techniques, such as the bootstrap, are also needed for the calculation of p-values. To avoid the computational burden of a nested bootstrap procedure, one may first evaluate $\sigma_0^i(\lambda)$ on a grid of parameter values and then approximate the required value $\sigma_0^i(\hat{\lambda})$ by interpolation. For those statistics T_n^i whose null distributions are independent of λ, the weights described in case (i) are retained.

Remark 5.1. *In case (ii), the weights are functions of $\hat{\lambda}_n$ and hence are random, so Theorem 1.12 is not directly applicable. However, if $\hat{\lambda}_n$ is a consistent estimator of λ under the null hypothesis and the mapping $\lambda \mapsto \sigma_0^i(\lambda)$ is continuous and strictly positive at the true parameter value, then*

$$\sigma_0^i(\hat{\lambda}_n) \xrightarrow[n\to\infty]{P} \sigma_0^i(\lambda).$$

Since the limiting distribution still depends on λ, practical implementation requires a bootstrap or another appropriate approximation method to obtain critical values and p-values.

5.2 Applications to different testing problems

This section has two main objectives. The first is to show how linear combinations of test statistics across different tuning parameters can serve as an alternative to data-dependent parameter selection. The second is to demonstrate how the same idea can be applied to combine different, yet complementary, classes of test statistics, thereby obtaining a compromise test with satisfactory performance against a broader class of alternatives.

To provide a basis for comparison with the proposed test, the data-dependent procedure of Tenreiro 2019 for selecting the tuning parameter is presented below.

Algorithm 4 Automatic selection of the smoothing parameter

1: Consider a finite set of candidate smoothing parameters $\Lambda = \{\lambda_1, \ldots, \lambda_K\}$.
2: Generate B independent samples of size n from the null distribution.
3: For each Monte Carlo sample $b = 1, \ldots, B$ and each $\lambda \in \Lambda$, compute the statistic $T_{n,\lambda}^{(b)}$.
4: Choose a regular grid $G_p = \{p, 2p, 3p, \ldots\} \cap (0, 1)$, where $0 < p \leq \alpha/K$.
5: For each $\lambda \in \Lambda$ and each $u \in G_p$, estimate the critical value $c_{n,\lambda}(u)$ as the empirical $(1-u)$-quantile of $T_{n,\lambda}^{(1)}, \ldots, T_{n,\lambda}^{(B)}$.
6: For each $u \in G_p$, estimate $\psi_\lambda(u) = \mathbb{P}\left(T_{n,\lambda_u} > c_{n,\lambda_u}(u)\right)$ by

$$\widehat{\psi}_{\bar{\lambda}}(u) = \frac{1}{B} \sum_{b=1}^{B} \mathbf{1}\left\{ T_{n,\bar{\lambda}_u^{(b)}}^{(b)} > c_{n,\bar{\lambda}_u^{(b)}}(u) \right\},$$

where $\bar{\lambda}_u^{(b)}$ is defined by

$$\bar{\lambda}_u^{(b)} = \arg\max_{\lambda \in \Lambda} \left(T_{n,\lambda}^{(b)} - c_{n,\lambda}(u) \right).$$

7: Choose

$$u_{n,\alpha,p}^{\bar{\lambda}} = \max\left\{ u \in G_p : \widehat{\psi}_{\bar{\lambda}}(u) \leq \alpha \right\}.$$

8: For the observed sample, compute $T_{n,\lambda}^{\text{obs}}$, $\lambda \in \Lambda$.
9: Select the smoothing parameter

$$\widehat{\lambda} = \arg\max_{\lambda \in \Lambda} \left(T_{n,\lambda}^{\text{obs}} - c_{n,\lambda}(u_{n,\alpha,p}^{\bar{\lambda}}) \right).$$

The flexibility of the proposed approaches is illustrated through simulation examples from linear and directional settings, covering real-valued univariate

and multivariate distributions, as well as circular and spherical distributions. Throughout the power study, the significance level is fixed at $\alpha = 0.05$.

5.2.1 Testing exponentiality

Let $X_1, \ldots, X_n$ be an i.i.d. sample from a non-negative random variable whose distribution function F is unknown. The problem considered here is testing whether the sample comes from an exponential distribution with density

$$f(x;\lambda) = \lambda e^{-\lambda x}, \quad x \geq 0,\ \lambda > 0,$$

namely

$$H_0 : X_1 \sim \mathcal{E}(\lambda) \quad \text{for some } \lambda > 0,$$

against the alternative that no exponential distribution provides the underlying model,

$$H_1 : X_1 \not\sim \mathcal{E}(\lambda) \quad \text{for all } \lambda > 0.$$

Testing exponentiality is a classical problem in gof testing, and a wide range of procedures has been developed for this purpose (see, e.g., Henze and Meintanis, 2005 and Torabi et al., 2018 for comprehensive reviews).

Among characterization-based procedures, the test proposed in Cuparić et al., 2019 is included as a relevant representative, since it has been reported to have favorable power properties. It is based on the Puri-Rubin characterization (Puri and Rubin, 1970), according to which if X_1 and X_2 are independent copies of a positive random variable, then

$$|X_1 - X_2| \stackrel{\mathcal{D}}{=} X_1$$

holds iff X_1 follows an exponential distribution. This characterization is particularly convenient for gof testing, since it transforms the problem of testing exponentiality into assessing whether the original observations and the absolute differences generated from independent pairs have the same distribution.

Let $Y_1, \ldots, Y_n$, denote the standardized sample, where

$$Y_i = \hat{\lambda}_n X_i, \quad i = 1, ..., n,$$

with $\hat{\lambda}_n = 1/\bar{X}$. This leads to a test statistic defined through a weighted L^2-type distance between the V-empirical Laplace transforms of the random quantities appearing in the characterization:

$$M_n^{(a)}(\hat{\lambda}) = \int_0^\infty \left[L_n^{(1)}(t) - L_n^{(2)}(t)\right]^2 e^{-at} dt.$$

Here, $a > 0$ is a tuning parameter, and

$$L_n^{(1)}(t) = \frac{1}{n}\sum_{i=1}^{n} e^{-tY_i},$$
$$L_n^{(2)}(t) = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} e^{-t|Y_i - Y_j|}$$

are the V-empirical Laplace transforms of Y_1 and $|Y_1 - Y_2|$, respectively,

Under the null hypothesis, the distribution of the test statistic is free of λ. Furthermore, it admits a representation as a degenerate V-statistic with an estimated parameter, and the corresponding symmetrized kernel is given by

$$\Phi(x_i, x_j, x_k, x_l; a) = \frac{1}{4!}\sum_{\pi\in\Pi(4)} \Big(\frac{1}{|x_{\pi(i)} - x_{\pi(j)}| + |x_{\pi(k)} - x_{\pi(l)}| + a} - \frac{2}{|x_{\pi(i)} - x_{\pi(j)}| + x_{\pi(k)} + a} + \frac{1}{x_{\pi(i)} + x_{\pi(j)} + a}\Big).$$

where $\Pi(4)$ is the set of all permutations of the set $\{i, j, k, l\}$. The second projection of this kernel therefore takes the form

$$\begin{aligned}\Phi_2(x_i, x_j, a) = &-\frac{1}{2} + \frac{1}{3}(e^{-x_i} + e^{-x_j}) + \frac{1}{6(a + x_i + x_j)} \\ &+ \frac{1}{6}e^{a - x_i - x_j}\mathrm{Ei}(-a)\left[a(e_i^x - 2)(e_j^x - 2) - e^{x_i} - e^{x_j} + 4\right] \\ &+ \frac{1}{6}e^{-a - x_i - x_j}\left[\mathrm{Ei}(a)(4a + e^{x_i} + e^{x_j} - 4) - \mathrm{Ei}(a + x_i)(4(a + x_i - 1) + e^{x_j})\right. \\ &\left.+\mathrm{Ei}(a + x_j)(4(a + x_j - 1) + e^{x_i}) - 4(a + x_i + x_j - 1)\mathrm{Ei}(a + x_i + x_j)\right],\end{aligned}$$

where $\mathrm{Ei}(x) = -\int_{-x}^{\infty} \frac{e^{-t}}{t} dt$ denotes the exponential integral.

An additional power study, which extended the settings considered in Cuparić et al., 2019, was conducted to identify suitable values of the tuning parameter for constructing the linear combination. The set of alternatives was taken

from Cuparić et al., 2022. In addition, for the purpose of an external evaluation of the proposed procedure, the log-logistic distribution $LL(\beta)$ with shape parameter $\beta > 0$ and scale parameter fixed at one, having density

$$f(x;\beta) = \frac{\beta x^{\beta-1}}{\left(1+x^{\beta}\right)^2}, \quad x > 0,$$

and the Gompertz distributions $Gomp(\mu, \theta)$ defined in (3.24) were also included. The sample size was fixed at $n = 50$, with $N = 10000$ replicates.

Dist.	$M_n^{(0.2)}$	$M_n^{(0.5)}$	$M_n^{(0.7)}$	$M_n^{(1)}$	$M_n^{(1.5)}$	$M_n^{(2)}$	$M_n^{(3)}$	$M_n^{(5)}$	$M_n^{(7)}$	$M_n^{(10)}$	$M_n^{(15)}$	$M_n^{(\hat{a}_{dd})}$	LC1	LC2
$\mathcal{E}(1)$	5	5	5	5	5	5	5	5	5	5	5	5	5	5
$W(1.4)$	81	84	84	86	86	86	87	86	87	86	86	85	87	87
$\Gamma(2)$	97	97	97	97	97	97	96	95	95	95	94	97	96	96
HN	42	48	50	52	55	58	61	62	62	63	64	53	61	61
U	91	95	96	97	98	98	99	99	99	99	100	99	99	99
$CH(0.5)$	90	90	91	90	91	91	91	91	91	91	91	92	92	92
$CH(1)$	28	33	36	37	39	41	46	45	46	47	47	43	46	46
$CH(1.5)$	99	100	100	100	100	100	100	100	100	100	100	100	100	100
$LF(2)$	58	64	67	68	71	74	77	77	77	78	78	73	76	76
$LF(4)$	77	82	85	86	88	90	91	92	91	91	92	89	91	91
$EV(1.5)$	74	81	84	85	89	90	91	93	93	94	94	91	92	92
$LN(0.8)$	98	94	93	89	85	82	79	72	67	64	61	95	78	78
$LN(1.5)$	23	35	42	48	58	64	74	80	84	87	89	87	76	76
$DL(1)$	82	77	76	72	69	67	66	59	58	55	53	76	64	65
$DL(1.5)$	100	100	100	100	100	100	100	99	99	99	99	100	99	99
$W(0.8)$	15	17	19	19	22	23	28	29	30	32	36	26	27	27
$\Gamma(0.4)$	98	97	97	97	97	97	98	97	98	98	98	97	98	98
$LL(0.8)$	8	16	17	23	28	36	44	58	66	74	79	69	52	51
$LL(1.5)$	94	90	87	86	85	87	87	91	94	100	100	98	91	90
$Gomp(0.5,1)$	29	33	35	37	40	42	43	44	47	46	47	38	44	44
$Gomp(1.5,1)$	74	81	84	86	88	90	91	92	93	94	94	91	92	92

Table 5.1: Empirical size and powers (in percentages) of the $M_n^{(a)}$ statistics and their linear combinations for $n = 50$

The results presented in Table 5.1 show that, for certain alternatives, the empirical power is strongly influenced by the choice of the tuning parameter a. Therefore, as a compromise procedure, a linear combination of $M_n^{(0.7)}$ and $M_n^{(7)}$ was considered. The corresponding weights were chosen to be inversely proportional to the exact and limiting null standard deviations of $nM_n^{(a)}$, leading to the test statistics denoted by LC1 and LC2 in Table 5.1, respectively.

The obtained results indicate that the proposed combined test has power close to that of the best-performing individual test. It is also worth noting that, although the exact and limiting null standard deviations differ, as shown in Table 5.2, this difference does not affect the empirical power of the combined tests.

Moreover, the comparison with the data-driven procedure of Tenreiro, 2019, denoted by $M_n^{(\hat{a}_{dd})}$ in Table 5.1, suggests that the proposed linear combination approach is competitive and, in many cases, leads to higher empirical power. At the same time, the results show that this advantage is not systematic, since the data-driven procedure performs better for some alternatives. From a computational

a	Exact ($n = 20$)	Exact ($n = 50$)	Asymptotic
0.2	$4.88 \cdot 10^{-1}$	$3.60 \cdot 10^{-1}$	$2.80 \cdot 10^{-1}$
0.5	$1.76 \cdot 10^{-1}$	$1.46 \cdot 10^{-1}$	$1.11 \cdot 10^{-1}$
0.7	$1.18 \cdot 10^{-1}$	$9.41 \cdot 10^{-2}$	$7.33 \cdot 10^{-2}$
1	$7.06 \cdot 10^{-2}$	$5.61 \cdot 10^{-2}$	$4.48 \cdot 10^{-2}$
1.5	$3.57 \cdot 10^{-2}$	$2.92 \cdot 10^{-2}$	$2.39 \cdot 10^{-2}$
2	$2.16 \cdot 10^{-2}$	$1.79 \cdot 10^{-2}$	$1.46 \cdot 10^{-2}$
3	$9.74 \cdot 10^{-3}$	$7.98 \cdot 10^{-3}$	$6.87 \cdot 10^{-3}$
5	$3.20 \cdot 10^{-3}$	$2.87 \cdot 10^{-3}$	$2.39 \cdot 10^{-3}$
7	$1.44 \cdot 10^{-3}$	$1.23 \cdot 10^{-3}$	$1.12 \cdot 10^{-3}$
10	$5.89 \cdot 10^{-4}$	$5.21 \cdot 10^{-4}$	$4.76 \cdot 10^{-4}$
15	$2.09 \cdot 10^{-4}$	$1.84 \cdot 10^{-4}$	$1.71 \cdot 10^{-4}$

Table 5.2: Standard deviations of the $nM_n^{(a)}$ statistics under the null hypothesis.

standpoint, the proposed approach has a clear practical advantage: it avoids the additional calibration step required by the data-driven selection rule. Consequently, its implementation is considerably simpler and less time-consuming in repeated simulation settings.

The results obtained for the additional alternatives lead to the same overall conclusions as those observed for the alternatives used in the initial selection step.

5.2.2 Testing multivariate normality

Let $X_1, \ldots, X_n$ be an i.i.d. sample from a d-variate random vector X. The composite hypothesis of interest is

$$H_0 : X \sim \mathcal{N}_d(\mu, \Sigma),$$

for some mean vector $\mu \in \mathbb{R}^d$ and $d \times d$ positive definite covariance matrix Σ. Multivariate normality testing has been investigated extensively, resulting in many affine invariant test procedures (a comprehensive account is given in Ebner and Henze, 2020). Many of the proposed tests depend on a tuning parameter, including the classical Baringhaus-Henze-Epps-Pulley (BHEP) test (Baringhaus and Henze, 1988; Henze and Zirkler, 1990). This has naturally raised the question of how to select the tuning parameter in practice.

The BHEP statistic is briefly reviewed below, along with the facts that will be used to construct the proposed procedure. Let $Y_1, \ldots, Y_n$ denote the standardized

sample, defined by

$$Y_i = \hat{\Sigma}_n^{-1/2}(X_i - \hat{\mu}_n), \quad i = 1, \ldots, n, \tag{5.1}$$

where $\hat{\mu}_n$ is the sample mean and $\hat{\Sigma}_n$ denotes the sample covariance matrix. For a tuning parameter $\beta > 0$, the BHEP test statistic is defined by

$$\text{BHEP}_n^{(\beta)} = n \int_{\mathbb{R}^d} \left| \varphi_n(t) - e^{-\|t\|^2/2} \right|^2 w_\beta(t)\, dt,$$

where

$$\varphi_n(t) = \frac{1}{n} \sum_{j=1}^{n} e^{it^\top Y_j}$$

is the empirical characteristic function of the standardized sample, while

$$w_\beta(t) = \frac{1}{(2\pi\beta^2)^{d/2}} \exp\left\{ -\frac{\|t\|^2}{2\beta^2} \right\}, \; t \in \mathbb{R}^d$$

is a Gaussian weight function.

By evaluating the integral explicitly, the statistic can be written in the form of a V-statistic with estimated parameters:

$$\text{BHEP}_n^{(\beta)} = \frac{1}{n} \sum_{j=1}^{n} \sum_{k=1}^{n} e^{-\frac{\beta^2 \|Y_j - Y_k\|^2}{2}} - \frac{2}{(1+\beta^2)^{d/2}} \sum_{j=1}^{n} e^{-\frac{\beta^2 \|Y_j\|^2}{2(1+\beta^2)}} + \frac{n}{(1+2\beta^2)^{d/2}}.$$

Then, their asymptotic variance is given by (Henze and Zirkler, 1990)

$$\begin{aligned} \text{Var}(\text{BHEP}_n^{(\beta)}) =& \frac{2}{(1+4\beta^2)^{d/2}} + \frac{2}{(1+2\beta^2)^d} \left[1 + \frac{8d\beta^4(1+2\beta^2)^2 + 3d(d+2)\beta^8}{4(1+2\beta^2)^4} \right] \\ & - \frac{4}{w(\beta)^{d/2}} \left[1 + \frac{3d\beta^4}{2q(\beta)} + \frac{d(d+2)\beta^8}{2q(\beta)^2} \right], \end{aligned} \tag{5.2}$$

where $q(\beta) = (1+\beta^2)(1+3\beta^2)$.

The role of the tuning parameter β in the BHEP test has been analyzed in detail in Tenreiro, 2009; Tenreiro, 2017. Based on these studies, the following choices are recommended:

$$\beta_S = \frac{\sqrt{2}}{0.896 + 0.052d}, \quad \text{for short-tailed alternatives,}$$

$$\beta_L = \frac{\sqrt{2}}{1.856 + 0.098d}, \quad \text{for long-tailed alternatives,}$$

$\beta_T = \dfrac{\sqrt{2}}{1.376 + 0.075d}$, when no prior information about the alternative is available.

In this chapter, instead of selecting a single value of β, several BHEP statistics are combined. Two linear combinations are considered. The first, denoted by LC1, combines the statistics corresponding to β_S and β_L. The second, denoted by LC2, uses the more widely separated values $\beta = 0.2$ and $\beta = 2$, which were chosen based on the observed power behavior in the simulation study. The weights are determined from the asymptotic variances in (5.2), since the finite-sample distribution of the BHEP statistic is known to rapidly approach its limiting distribution.

A power study was carried out for the sample size $n = 50$ with $N = 50000$ replicates. The alternatives used for the initial assessment were taken from Ebner and Henze, 2020, whereas d-variate Laplace (La^d), logistic (Lo^d), and skew-normal (SN^d) distributions with independent components were included as additional alternatives for an external evaluation of the proposed procedure. The results obtained for dimensions $d = 2$ and $d = 5$ are reported in Tables 5.3 and 5.4, respectively.

When no prior information about the form of the alternative is available, LC1 and LC2 exhibit performance very close to that of the recommended statistic based on the tuning parameter β_T. In comparison with the data-driven selection method (Tenreiro, 2019), which powers are presented in column $\beta_{T,\hat{a}}$, the proposed linear combination approach is either competitive or more powerful, both for the alternatives used in the initial study and for the additional alternatives considered for external evaluation.

The relatively low powers observed for the uniform $U^d(0,1)$ and beta $B^d(2,2)$ alternatives are mainly caused by the zero power of the component corresponding to $\beta = 0.2$. However, this effect becomes less pronounced for larger sample sizes. For example, in dimension $d = 5$ and for $n = 300$, the powers of LC2 for these two alternatives are equal to 100 and 28, respectively, although one of its components still has zero empirical power.

Dist.	β_S	β_L	β_T	$\beta_{0.2}$	$\beta_{0.5}$	β_1	β_2	β_5	β_{10}	$\beta_{T,\hat{a}}$	LC1	LC2
$N_d(\mathbf{0}, I_d)$	5	5	5	5	5	5	5	5	5	5	5	5
$NMIX(0.5, \mathbf{3}, I_d)$	83	21	56	2	6	65	85	59	28	68	64	52
$NMIX(0.7, \mathbf{3}, I_d)$	91	69	84	25	48	87	90	60	29	77	94	77
$NMIX(0.9, \mathbf{3}, I_d)$	73	87	83	86	88	82	64	30	16	83	84	86
$NMIX(0.5, \mathbf{0}, B_d)$	31	28	30	21	25	31	29	17	11	27	31	30
$NMIX(0.9, \mathbf{0}, B_d)$	29	44	38	44	46	37	24	12	9	42	40	43
$t_3(\mathbf{0}, I_d)$	79	84	83	79	83	83	74	54	35	81	84	84
$t_5(\mathbf{0}, I_d)$	42	53	50	52	54	48	36	19	13	47	50	52
$t_{10}(\mathbf{0}, I_d)$	14	21	18	24	23	17	11	8	7	20	18	22
$U^d(0,1)$	67	22	55	0	3	60	65	38	20	48	55	25
$LN^d(0, 0.5)$	93	98	97	98	98	97	88	59	33	96	97	98
$B^d(1,2)$	79	74	80	36	60	81	73	47	24	66	79	69
$B^d(2,2)$	17	4	11	0	1	13	17	9	6	9	11	4
$P^d_{II}(0.5, 0, 1)$	100	88	99	0	13	99	100	99	91	99	99	94
$P^d_{VII}(5, 0, 1)$	33	44	40	44	45	39	27	15	11	39	40	43
$P^d_{VII}(10, 0, 1)$	11	17	14	20	18	13	9	7	6	16	15	17
$La^d(0,1)$	61	63	64	51	59	64	57	37	22	59	65	64
$Lo^d(0,1)$	15	22	20	24	24	19	12	8	7	20	20	23
$SN^d(0,1,2)$	17	24	22	25	26	21	13	8	6	19	21	23
$SN^d(0,1,5)$	64	78	75	72	78	73	53	26	14	67	75	75

Table 5.3: Empirical size and powers (in percentages) for the $BHEP$ statistics and their linear combinations LC1 and LC2 for $n = 50$ and $d = 2$

Dist.	β_S	β_L	β_T	$\beta_{0.2}$	$\beta_{0.5}$	β_1	β_2	β_5	β_{10}	$\beta_{T,\hat{a}}$	LC1	LC2
$N_d(\mathbf{0}, I_d)$	5	5	5	5	5	5	5	5	5	5	5	5
$NMIX(0.5, \mathbf{3}, I_d)$	52	11	24	3	6	41	47	10	9	33	30	21
$NMIX(0.7, \mathbf{3}, I_d)$	72	40	57	13	30	68	54	11	9	51	84	40
$NMIX(0.9, \mathbf{3}, I_d)$	75	95	92	88	95	86	37	10	9	92	92	90
$NMIX(0.5, \mathbf{0}, B_d)$	99	96	98	78	93	98	96	33	19	97	98	97
$NMIX(0.9, \mathbf{0}, B_d)$	67	90	83	94	92	76	43	12	10	89	85	92
$t_3(\mathbf{0}, I_d)$	97	99	99	99	99	98	92	43	24	99	99	99
$t_5(\mathbf{0}, I_d)$	72	86	82	87	87	78	51	15	11	84	83	86
$t_{10}(\mathbf{0}, I_d)$	22	42	33	49	45	28	15	7	7	41	36	45
$U^d(0,1)$	50	8	33	0	2	48	30	6	7	28	31	3
$LN^d(0, 0.5)$	96	100	99	100	100	99	78	19	13	99	99	99
$B^d(1,2)$	67	64	73	21	53	73	40	9	8	52	72	42
$B^d(2,2)$	15	3	8	0	1	14	9	4	5	5	8	1
$P^d_{II(0.5,0,}1)$	97	35	88	0	6	96	89	22	22	91	89	31
$P^d_{VII}(5, 0, 1)$	33	55	46	61	58	40	20	8	7	56	48	57
$P^d_{VII}(10, 0, 1)$	10	19	14	24	21	12	8	6	6	18	14	19
$La^d(0,1)$	61	73	70	70	73	65	44	14	10	72	71	73
$Lo^d(0,1)$	13	25	19	31	27	15	9	6	6	22	20	26
$SN^d(0,1,2)$	13	24	20	25	25	16	8	6	6	18	20	22
$SN^d(0,1,5)$	58	83	77	77	83	69	31	8	7	71	77	75

Table 5.4: Empirical size and powers (in percentages) for the $BHEP$ statistics and their linear combinations LC1 and LC2 for $n = 50$ and $d = 5$

The results reported in Ebner and Henze, 2020 indicate that the spherical-harmonics-based test proposed in Manzotti and Quiroz, 2001 exhibits power properties that are complementary to those of the BHEP test. This observation suggests that a combination of the two procedures may yield a test with enhanced sensitivity against a broader class of alternatives.

Let $\mathcal{H}_j$ denote a fixed orthonormal basis of the space of spherical harmonics of degree j on S^{d-1} with respect to the uniform measure, and define

$$\mathcal{G}_j = \bigcup_{k=0}^{j} \mathcal{H}_k.$$

Furthermore, let

$$r_j(x) = \|x\|^j, \text{ for } x \in \mathbb{R}^d \text{ and } u(x) = \frac{x}{\|x\|}, \ x \neq 0.$$

For a fixed dimension d, consider the following collection of functions:

$$f_1 = r_1, \qquad \{f_2, \ldots, f_k\} = \{r_3(g \circ u) : g \in \mathcal{G}_2\}.$$

Since $\mathcal{G}_2$ contains $\binom{d+1}{2} + d$ elements, the collection consists of

$$k = \binom{d+1}{2} + d + 1$$

functions. The corresponding vector-valued function is denoted by

$$\mathbf{f} = (f_1, \ldots, f_k)^\top.$$

Let $Z \sim \mathcal{N}_d(0, I_d)$ be a standard d-variate normal random vector and let $Y_1, \ldots, Y_n$ denote the standardized sample defined in (5.1). Define the $k \times k$ matrix $V = (v_{ij})_{i,j=1}^k$ by

$$v_{ij} = \mathbb{E}\left[f_i(Z) f_j(Z)\right] - \mathbb{E} f_i(Z)\, \mathbb{E} f_j(Z), \quad i, j = 1, \ldots, k.$$

Assuming that V is nonsingular, set

$$v_n(f_j) = \frac{1}{\sqrt{n}} \sum_{\ell=1}^{n} \left\{f_j(Y_\ell) - \mathbb{E} f_j(Z)\right\}, \quad j = 1, \ldots, k,$$

and collect these quantities in the vector

$$\mathbf{v}_n(\mathbf{f}) = \big(v_n(f_1), \ldots, v_n(f_k)\big)^\top.$$

The spherical-harmonics-based test statistic is then defined as

$$MQ_n^{(2)}(\mathbf{f}) = \mathbf{v}_n(\mathbf{f})^\top V^{-1} \mathbf{v}_n(\mathbf{f}).$$

This statistic admits a V-statistic representation with a kernel

$$\Phi(x_1, x_2; \mu, \Sigma) = \sum_{i=1}^{k}\sum_{j=1}^{k} m_{ij} \left[f_i\Big(\Sigma^{-1/2}(x_1-\mu)\Big) - \mathbb{E} f_i\Big(\Sigma^{-1/2}(X-\mu)\Big)\right] \times \left[f_j\Big(\Sigma^{-1/2}(x_2-\mu)\Big) - \mathbb{E} f_j\Big(\Sigma^{-1/2}(X-\mu)\Big)\right],$$

where m_{ij} denotes the (i,j)-th entry of the matrix V^{-1}.

The asymptotic variance of this test statistic can be expressed in terms of the eigenvalues given in Manzotti and Quiroz, 2001. Since the smallest eigenvalue has to be computed numerically and its contribution is negligible, the variance is approximated by retaining the first $k-1$ largest eigenvalues:

$$\mathrm{Var}\Big[MQ_n^{(2)}(\mathbf{f})\Big] \approx 2\left[d\left(\frac{2}{d+4}\right)^2 + \frac{d^2+d-2}{2}\left(1 - \frac{2(d+1)^2(d+3)^2\Gamma\left(\frac{d+1}{2}\right)^2}{d^2(d+2)^2(d+4)\Gamma\left(\frac{d}{2}\right)^2}\right)^2 + 1\right].$$

A linear combination is then formed from the BHEP statistic with tuning parameter β_T and the statistic $MQ_n^{(2)}$. As before, the weights are chosen to be inversely proportional to the corresponding asymptotic null standard deviations.

The power study was carried out under the same settings as those used for the BHEP test. The results, reported in Table 5.5, indicate that the power of the combined procedure is generally close to the larger of the two component powers, and in some cases even exceeds both individual tests. The same overall conclusion is obtained for the alternatives included in the external evaluation.

Dist.	$d=2$ $BHEP$	$MQ^{(2)}$	LC	$d=5$ $BHEP$	$MQ^{(2)}$	LC
$\mathrm{N}_d(\mathbf{0}, I_d)$	5	5	5	5	5	5
$\mathrm{NMIX}(0.5, \mathbf{3}, I_d)$	57	51	65	45	48	58
$\mathrm{NMIX}(0.7, \mathbf{3}, I_d)$	93	22	83	85	18	76
$\mathrm{NMIX}(0.9, \mathbf{3}, I_d)$	84	62	84	79	55	80
$\mathrm{NMIX}(0.5, \mathbf{0}, B_d)$	31	31	35	28	27	32
$\mathrm{NMIX}(0.9, \mathbf{0}, B_d)$	39	50	50	35	45	45
$t_3(\mathbf{0}, I_d)$	84	91	91	78	88	88
$t_5(\mathbf{0}, I_d)$	49	66	64	42	60	59
$t_{10}(\mathbf{0}, I_d)$	18	29	27	15	25	23
$U^d(0,1)$	56	83	80	48	76	73
$LN^d(0,0.5)$	97	83	96	95	78	93
$B^d(1,2)$	80	40	77	75	35	72
$B^d(2,2)$	12	34	26	10	30	22
$P^d_{II}(0.5,0.1)$	99	100	100	98	100	100
$P^d_{VII}(5,0.1)$	40	55	53	35	50	48
$P^d_{VII}(10,0.1)$	14	22	21	12	19	18
$La^d(0,1)$	64	78	76	69	83	84
$Lo^d(0,1)$	20	30	29	20	30	30
$SN^d(0,1,2)$	22	13	19	20	11	17
$SN^d(0,1,5)$	75	34	68	77	31	69

Table 5.5: Empirical size and powers (in percentages) for the $BHEP$, $MQ2$ and their linear combination LC, for $d=2$ and $d=5$

5.2.3 Testing wrapped Cauchy distribution on the circle

Let $\Theta_1, \ldots, \Theta_n$ be an i.i.d. sample from a distribution defined on a circle S^1. The null hypothesis of interest is

$$H_0 : \Theta_1 \sim \mathcal{WC}(\mu, \rho),$$

where $\mathcal{WC}(\mu, \rho)$ stands for wrapped Cauchy distribution with density

$$f(\theta; \mu, \rho) = \frac{1}{2\pi} \frac{1 - \rho^2}{1 + \rho^2 - 2\rho \cos(\theta - \mu)}, \quad \theta \in [-\pi, \pi),\ \mu \in [-\pi, \pi),\ \rho \in [0, 1).$$

This unimodal and symmetric distribution was introduced by Lévy, 1939 and later investigated by Wintner, 1947. Its relevance in circular statistics is further supported by the result of McCullagh, 1996, which showed that this family is closed under conformal maps that preserve the unit circle, known as Möbius transformations.

In Jammalamadaka et al., 2019, a powerful class of gof tests for circular distributions is proposed. For testing the wrapped Cauchy distribution, and with the weights chosen according to a Poisson law, the corresponding test statistic is given by

$$C_{n,\lambda} = n \sum_{r=0}^{\infty} \left| \varphi_n(r) - \varphi(r; 0, \hat{\rho}_n) \right|^2 \cdot \frac{\lambda^r}{r!} e^{-\lambda},$$

where φ_n denotes the empirical characteristic function of the centered observations $\theta_i - \hat{\mu}_n$, while

$$\varphi(r; \mu, \rho) = \rho^{|r|} e^{ir\mu}, \ \ r \in \mathbb{Z},$$

is the characteristic function of the wrapped Cauchy distribution. The quantities $\hat{\mu}_n$ and $\hat{\rho}_n$ are the mle of μ and ρ, respectively. Centering the observations by $\hat{\mu}_n$ removes the dependence on the location parameter so that the null distribution may be studied under $\mu = 0$ without loss of generality.

It can be verified that this statistic admits a V-statistic representation with estimated parameters whose kernel is equal to

$$\Phi(\theta_1, \theta_2; \mu, \rho, \lambda) = \sum_{r=0}^{\infty} \Big(\cos(r(\theta_1 - \theta_2)) - \rho^{|r|} \big(\cos(r(\theta_1 - \mu)) + \cos(r(\theta_2 - \mu)) \big) + \rho^{2|r|} \Big) e^{-\lambda} \lambda^r / r!.$$

In contrast to the examples previously considered, both the finite-sample and the limiting distribution under the null hypothesis depend on the parameter ρ.

Hence, in order to apply the proposed linear combination approach, the variance function is approximated, for each selected sample size n, by the following procedure:

- A grid $G_\rho = \{\rho_1, \ldots, \rho_{100}\}$ of 100 equally spaced values is constructed on the interval $[0, 1)$, where

$$\rho_j = \frac{j-1}{100}, \quad j = 1, \ldots, 100.$$

- For each value of ρ_j in the grid, $N = 500$ samples of size $n = 50$ are generated from the wrapped Cauchy distribution centered at zero with concentration parameter ρ_j, and the empirical variance of the test statistic is computed.

- The empirical variances computed on the selected parameter grid are approximated by linear interpolation using the function `approxfun` from the R package `stats` (R Core Team, 2022). This function constructs an interpolation function from the supplied pairs of grid points and empirical variances, allowing evaluation at arbitrary parameter values lying between the grid points.

The bootstrap procedure described in Jammalamadaka et al., 2019 is then applied, using the plug-in approximation of the variance obtained from the interpolation step. The case $n = 50$ is considered, with $N = 10000$ replications. The proposed linear combination is constructed from the statistics corresponding to $\lambda = 0.3$ and $\lambda = 1$. The results are reported in Table 5.6.

In addition to the alternatives already introduced in Chapter 4.3, the following distributions and mixture alternatives are considered:

- a cardioid distribution $\mathcal{C}a(\mu, \rho)$

$$f(\theta) = \frac{1}{2\pi}(1 + 2\rho\cos(\theta - \mu)), \ -\pi \leq \theta, \mu < \pi, \ |\rho| < \frac{1}{2};$$

- a wrapped normal distribution $\mathcal{WN}(\mu, \sigma)$

$$f(\theta) = \frac{1}{\sigma\sqrt{2\pi}} \sum_{m=-\infty}^{\infty} exp\Big[\frac{-(\theta - \mu + 2\pi m)^2}{2\sigma^2}\Big], \ -\pi \leq \theta, \mu < \pi, \ \sigma > 0;$$

- a wrapped Laplace distribution $\mathcal{WL}a(\mu, \sigma)$

$$f(\theta) = \frac{1}{2\sigma\left(1 - e^{-2\pi/\sigma}\right)}\left(e^{-|\theta-\mu|/\sigma} + e^{-(2\pi-|\theta-\mu|)/\sigma}\right),\ -\pi \leq \theta, \mu < \pi,\ \sigma > 0;$$

- a mixture of two wrapped Cauchy distributions

$$p\mathcal{WC}(\mu_1, \rho_1) + (1-p)\mathcal{WC}(\mu_2, \rho_2),$$

with mixture parameter $p \in (0, 1)$. In the simulations, the following notation is used.

Notation	Mixture
$\mathcal{WC}_{M1}$	$0.6\text{WC}(0, 0.8) + 0.4\text{WC}(\frac{\pi}{4}, 0.5)$
$\mathcal{WC}_{M2}$	$0.5\text{WC}(0, 0.5) + 0.5\text{WC}(\frac{\pi}{2}, 0.5)$
$\mathcal{WC}_{M3}$	$0.7\text{WC}(0, 0.9) + 0.3\text{WC}(\frac{\pi}{2}, 0.3)$
$\mathcal{WC}_{M4}$	$0.5\text{WC}(\frac{\pi}{2} - 0.3, 0.9) + 0.5\text{WC}(\frac{\pi}{2} + 0.3, 0.9)$
$\mathcal{WC}_{M5}$	$0.5\text{WC}(0, 0.2) + 0.5\text{WC}(-\pi, 0.2)$
$\mathcal{WC}_{M6}$	$0.9\text{WC}(0, 0.95) + 0.1\text{WC}(-\pi, 0.1)$
$\mathcal{WC}_{M7}$	$0.8\text{WC}(-\pi, 0.1) + 0.2\text{WC}(0, 0.99)$

As in the previous cases, the test statistics based on linear combinations generally perform better than the original test statistics for most values of λ and for almost all alternatives considered. A similar pattern is observed for the additional alternatives introduced in this section, suggesting that the improvement is not restricted to the distributions used in the initial comparison.

5.3 Combining test statistics of other types

Although the preceding developments were formulated within the framework of degenerate V-statistics, an analogous approach can be employed in more general settings. The present section applies the same principle to a test statistic that depends on a tuning parameter but does not belong to this class.

5.3.1 Testing uniformity on the sphere

Let $U_1, \ldots, U_n$ be an i.i.d. sample from a random vector U taking values on the unit sphere S^{d-1}. The problem of interest is to test whether the underlying distri-

Dist.	$C_{n,0.3}$	$C_{n,0.5}$	$C_{n,1}$	$C_{n,2}$	LC
$\mathcal{WC}(0.9)$	5	5	5	5	5
$\mathcal{WC}(0.5)$	5	5	5	5	5
$\mathcal{WC}(0.3)$	5	5	5	5	5
$\mathcal{VM}(0.5)$	5	5	5	5	5
$\mathcal{VM}(1)$	10	11	11	10	10
$\mathcal{VM}(2)$	45	44	41	34	43
$\mathcal{VM}(5)$	90	88	84	71	89
$\mathcal{VM}(7)$	91	91	89	78	93
$\mathcal{C}a(0.3)$	10	11	11	10	10
$\mathcal{C}a(0.5)$	56	58	56	43	56
$\mathcal{CW}(0.5)$	82	82	77	62	80
$\mathcal{CW}(1)$	56	58	56	43	56
$\mathcal{JP}(2,1)$	47	49	47	38	47
$\mathcal{JP}(2,0)$	46	44	40	33	42
$\mathcal{JP}(2,1.5)$	37	39	39	31	38
$\mathcal{B}a(3,0.5)$	65	64	61	48	63
$\mathcal{B}a(3,1)$	56	57	60	52	57
$\mathcal{B}a(5,1)$	70	76	83	81	83
$\mathcal{B}a(5,0.5)$	83	85	85	76	86
$\mathcal{WC}_{M1}$	28	26	25	24	25
$\mathcal{WC}_{M2}$	18	18	20	22	19
$\mathcal{WC}_{M3}$	91	91	91	86	90
$\mathcal{WC}_{M4}$	80	84	89	90	82
$\mathcal{WC}_{M5}$	6	6	7	7	6
$\mathcal{WC}_{M6}$	62	62	59	49	63
$\mathcal{WC}_{M7}$	39	44	51	65	47
$\mathcal{WL}a(0.2)$	28	30	32	27	32
$\mathcal{WL}a(0.5)$	24	23	20	15	22
$\mathcal{WL}a(0.8)$	10	9	9	8	10
$\mathcal{WN}(0.2)$	5	6	7	6	6
$\mathcal{WN}(0.5)$	32	33	33	27	33
$\mathcal{WN}(0.8)$	89	88	81	68	86

Table 5.6: Empirical powers (in percentages) of the $C_{n,\lambda}$ statistics and the linear combination LC based on $\lambda = 0.3$ and $\lambda = 1$, for sample size $n = 50$.

bution is uniform on the sphere, that is,

$$H_0 : U \sim \mathcal{U}(S^{d-1}).$$

Under H_0, the density of U with respect to the surface measure on S^{d-1} is given by

$$f(u) = \frac{1}{\omega_{d-1}}, \quad u \in S^{d-1},$$

where

$$\omega_{d-1} = \frac{2\pi^{d/2}}{\Gamma(d/2)}$$

denotes the surface area of S^{d-1}.

Uniformity testing on the hypersphere has been the subject of extensive research. The review paper by García-Portugués and Verdebout, 2018 gives a detailed overview of the main testing procedures, while simulation-based comparisons and empirical findings for particular procedures are reported in Diggle et al., 1985 and Figueiredo, 2007. In Borodavka and Ebner, 2026, a new class of test statistics $T_{n,\beta}$, based on powers of maximal projections, is proposed. The analysis below focuses on this class of tests.

Under the null hypothesis, by the rotational invariance of the uniform distribution on S^{d-1} and symmetry, the moment

$$\psi_d(\beta) = \mathbb{E}(b^\top U)^\beta$$

does not depend on the particular choice of $b \in S^{d-1}$. Moreover, for every $\beta \in \mathbb{N}$,

$$\psi_d(\beta) = \begin{cases} \dfrac{\Gamma\big((\beta+1)/2\big)\,\Gamma(d/2)}{\sqrt{\pi}\,\Gamma\big((\beta+d)/2\big)}, & \text{if } \beta \text{ is even,} \\ 0, & \text{if } \beta \text{ is odd,} \end{cases}$$

where $\Gamma(\cdot)$ denotes the gamma function. Based on this, the following family of test statistics $T_{n,\beta}$ was proposed:

$$T_{n,\beta} = T_{n,\beta}(U_1, \ldots, U_n) = n \max_{b \in S^{d-1}} \left(\frac{1}{n} \sum_{j=1}^{n} (b^\top U_j)^\beta - \psi_d(\beta) \right)^2, \quad \beta \in \mathbb{N}.$$

Motivated by the different sensitivity of $T_{n,\beta}$ to the choice of β, the linear-combination approach is applied to the statistics corresponding to $\beta = 5$ and $\beta = 6$, which exhibit complementary power properties. Since the null hypothesis

is simple, the weights are chosen as the inverses of the null standard deviations of the corresponding statistics, with these standard deviations approximated by Monte Carlo simulation based on $N_1 = 10000$ replicates. The alternatives are taken from Borodavka and Ebner, 2026. For the evaluation stage, the Watson distribution is considered. The Watson distribution on the unit sphere S^{d-1} is determined by an axial parameter $\mu \in S^{d-1}$ and a concentration parameter $\kappa \in \mathbb{R}$. Its density has the form

$$f(x;\mu,\kappa) = \frac{1}{M\left(\frac{1}{2},\frac{p}{2},\kappa\right)} \exp\left\{\kappa(\mu^\top x)^2\right\}, \quad x \in S^{p-1},$$

where $M(a,b,z)$ denotes Kummer's confluent hypergeometric function of the first kind, defined by the series

$$M(a,b,z) = \sum_{j=0}^{\infty} \frac{(a)_j}{(b)_j} \frac{z^j}{j!}.$$

Here, $(a)_j$ denotes the Pochhammer symbol, given by

$$(a)_j = a(a+1)\cdots(a+j-1), \quad j \geq 1,$$

with $(a)_0 = 1$. In Tables 5.7 and 5.8, $\mathrm{Wa}_1(\kappa)$ denotes the Watson distribution with mean direction $\mu = (1,0,\ldots,0)$, whereas $\mathrm{Wa}_2(\kappa)$ denotes the Watson distribution with mean direction given by the normalized vector $(1,\ldots,1)$. Both distributions are considered with the concentration parameter κ. The simulation study considers the cases $d = 2$ and $d = 3$. The sample size is fixed at $n = 100$, and empirical powers are estimated from $N_2 = 5000$ Monte Carlo replications.

The results reported in Tables 5.7 and 5.8 show that the power of the linear combination is consistently close to that of the more powerful component statistic, and in some cases even exceeds the power of both individual components. The results obtained for the additional alternatives confirm the same overall pattern.

Dist.	$T_{n,1}$	$T_{n,2}$	$T_{n,3}$	$T_{n,4}$	$T_{n,5}$	$T_{n,6}$	LC
$\mathcal{U}(S^{d-1})$	5	5	5	5	5	5	5
$\text{vMF}_1(0.05)$	6	5	6	5	6	5	5
$\text{vMF}_1(0.1)$	9	5	8	6	8	6	8
$\text{vMF}_1(0.25)$	32	6	31	5	28	5	21
$\text{vMF}_1(0.5)$	88	6	85	6	82	6	71
$\text{vMF}_1(0.75)$	100	11	100	12	99	12	98
$\text{vMF}_1(1)$	100	24	100	23	100	23	100
$\text{vMF}_1(2)$	100	98	100	98	100	98	100
$\text{Mix-vMF}_1(0.25)$	79	23	76	24	74	23	71
$\text{Mix-vMF}_1(0.5)$	4	23	5	24	5	23	16
$\text{Mix-vMF}_2(0.5)$	70	100	78	99	79	99	99
$\text{Mix-vMF}_2(0.75)$	30	82	28	80	27	79	80
$\text{Mix-vMF}_3(0.25)$	40	86	55	86	59	84	90
$\text{Mix-vMF}_4(0.33)$	11	72	21	71	28	69	75
$\text{Bing}_1(0.25)$	5	11	5	11	5	10	9
$\text{Bing}_1(0.5)$	5	33	5	32	5	30	23
$\text{Bing}_1(1)$	6	88	6	88	6	87	75
$\text{Bing}_2(0.1)$	5	23	5	22	5	21	16
$\text{Bing}_2(0.25)$	6	88	6	88	6	87	75
$\text{LP}_3(0.1)$	5	5	6	5	7	5	6
$\text{LP}_4(0.1)$	5	5	5	6	5	6	6
$\text{LP}_3(0.5)$	5	4	25	5	45	5	33
$\text{LP}_4(0.5)$	5	6	6	18	6	31	22
$\text{LP}_3(1)$	5	4	91	4	100	5	97
$\text{LP}_4(1)$	5	6	6	59	8	96	81
$\text{Wa}_1(0.5)$	5	34	5	34	5	33	26
$\text{Wa}_1(0.8)$	4	70	5	69	6	68	56
$\text{Wa}_2(\text{-}0.5)$	5	33	6	32	5	32	24
$\text{Wa}_2(\text{-}1)$	5	89	5	88	6	86	75
$\text{Wa}_2(\text{-}2)$	5	100	7	100	8	100	100

Table 5.7: Empirical power (in percentage) of the spherical uniformity tests and the linear combination based on $\beta = 5$ and $\beta = 6$, for $n = 100$ and $d = 2$

Dist.	$T_{n,1}$	$T_{n,2}$	$T_{n,3}$	$T_{n,4}$	$T_{n,5}$	$T_{n,6}$	LC
$\mathcal{U}(S^{d-1})$	5	5	5	5	5	5	5
$vMF_1(0.05)$	5	5	6	5	6	5	6
$vMF_1(0.1)$	6	5	7	5	7	5	6
$vMF_1(0.25)$	19	5	18	5	16	5	13
$vMF_1(0.5)$	65	5	58	6	52	5	40
$vMF_1(0.75)$	96	8	93	8	89	7	78
$vMF_1(1)$	100	14	100	14	99	13	97
$vMF_1(2)$	100	93	100	91	100	88	100
Mix-$vMF_1(0.25)$	63	15	58	15	50	14	46
Mix-$vMF_1(0.5)$	5	15	5	15	5	14	11
Mix-$vMF_2(0.5)$	85	99	92	99	92	99	100
Mix-$vMF_2(0.75)$	9	76	10	75	12	72	65
Mix-$vMF_3(0.25)$	62	90	73	88	75	82	93
Mix-$vMF_4(0.33)$	9	90	22	86	33	80	82
$Bing_1(0.25)$	5	14	5	13	6	12	11
$Bing_1(0.5)$	5	45	6	43	6	38	28
$Bing_1(1)$	6	98	7	98	9	96	89
$Bing_2(0.1)$	5	18	6	17	6	15	13
$Bing_2(0.25)$	6	83	7	80	8	75	63
$LP_3(0.1)$	5	5	6	5	5	5	5
$LP_4(0.1)$	6	5	6	5	5	5	5
$LP_3(0.5)$	5	4	7	5	10	5	9
$LP_4(0.5)$	5	5	5	6	5	7	7
$LP_3(1)$	5	5	19	5	34	5	24
$LP_4(1)$	5	5	5	9	6	17	13
$Wa_1(0.5)$	5	17	5	18	6	15	12
$Wa_1(0.8)$	6	46	6	45	6	38	30
$Wa_2(-0.5)$	6	14	6	13	6	12	9
$Wa_2(-1)$	5	51	6	45	6	38	24
$Wa_2(-2)$	7	99	8	97	10	93	80

Table 5.8: Empirical power (%) of the spherical uniformity tests and the linear combination based on $\beta = 5$ and $\beta = 6$, for $n = 100$ and $d = 3$

Chapter 6
Real data analysis

This chapter presents an analysis of several real data sets in order to illustrate the practical performance of the newly proposed tests. The procedures are applied to observational data arising from different settings, with the aim of testing the corresponding null hypotheses and assessing the adequacy of the considered probabilistic models. In this way, the empirical study complements the theoretical and simulation results given in the previous chapters and provides additional insight into the behavior of the proposed methods in real-data applications.

6.1 Applications to discrete data

In this section, the tests introduced in Chapter 2, together with the competing procedures discussed there, are applied to several real-world datasets in order to examine their practical performance in testing gof to the geometric distribution. The datasets considered come from a variety of application areas. Two of them come from capture-recapture studies, which are widely used in ecological and population estimation research; one dataset concerns the statistical modeling of natural catastrophes and is therefore relevant in the study of extreme events; and the final dataset relates to the distribution of plant and animal occurrences, providing an additional illustration in an ecological context. These datasets have been sourced from Anan et al., 2019, Bliss and Fisher, 1953 and Kagan, 2010.

1. *Bowel cancer data.* Screening for bowel cancer is carried out through a series of self-administered tests performed over the course of six consecutive days, each designed to detect the presence of blood in fecal samples. Individuals who have negative results on all six tests are not required to undergo any further medical procedures, while those with at least one positive test are subject to additional diagnostic steps. In this particular study, the recorded frequencies for positive results across the six days

were $(n_1, \ldots, n_6) = (8, 12, 16, 21, 12, 31)$, summing to a total sample size of $n = 100$.

2. *Dicentric data.* The dicentric chromosome dataset comprises measurements taken from cells exposed to a radiation dose of 0.405 Gy. In this experiment, the frequencies of cells exhibiting one to four dicentric chromosomes were observed as $(n_1, \ldots, n_4) = (66, 15, 1, 1)$, yielding a total of $n = 83$ analyzed cells.

3. *Primula data.* In plant ecology, quadrat counts are commonly used to assess the distribution of plant species, particularly their tendency to occur in distinct "clumps" rather than being evenly spread. This dataset contains records of the number of quadrats in which Primula auricula, a species of flowering plant from the Primulaceae family, was found. The observed frequency distribution is $(n_1, \ldots, n_9) = (21, 23, 14, 11, 4, 5, 4, 0, 1)$, where each n_k indicates the number of quadrats in which exactly k clumps of Primula auricula were found. The total number of quadrats surveyed is $n = 83$.

4. *Earthquake data.* The number of aftershocks recorded during the first 24 hours within a circular area of radius R (where R is determined by the magnitude of the main shock) is given by the observed frequency distribution $(n_0, \ldots, n_{34}) = (9, 6, 5, 4, 3, 1, 4, 3, 6, 4, 1, 0, 0, 0, 2, 0, 0, 2, 1, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1)$. This means that, for each possible number of aftershocks from 0 up to 34, the corresponding value in the sequence represents how many times that specific number of aftershocks was observed in the sample. The total number of observations in this dataset is $n = 55$.

The tests considered in Chapter 2 are formulated for the geometric distribution on the support $\{1, 2, \ldots\}$. Some of the real data sets considered above are zero-truncated count data, as is common in capture-recapture applications. If an underlying count variable X, defined on $\{0, 1, 2, \ldots\}$, has a geometric distribution, then, by the memoryless property, the conditional distribution of $X|X \geq 1$ is geometric on $\{1, 2, \ldots\}$. Therefore, the tests can be applied directly to the observed positive counts.

Based on that, the bowel cancer, dicentric and primula data sets are treated in this zero-truncated framework. Thus, only positive observations are used, and the frequency of zero, even when available, is treated as unknown. This is consistent with the support $\{1, 2, \ldots\}$ assumed in the definition of the tests. For earthquake data set, observations are recorded from zero and the zero frequency is

retained in the analysis. Hence, this data set is not treated as zero-truncated. To apply tests defined on $\{1, 2, \ldots\}$, the observations are shifted by one.

Table 6.1 presents the p-values calculated using the parametric bootstrap method with $B = 1000$ bootstrap samples. For the first real-life data set, all tests agree in rejecting H_0. In contrast, for the second and fourth data sets, none of the tests provide evidence against H_0, which aligns with the findings reported by Anan et al., 2019 and Kagan, 2010. For the third data set, the tests K_n, W_n, and H_n do not indicate evidence against H_0, whereas the remaining tests, including the newly proposed ones, reject the null hypothesis. Since a satisfactory fit of the negative binomial distribution to this dataset was reported in Bliss and Fisher, 1953, the obtained results are not without support. At the same time, because the geometric distribution represents a more restrictive submodel of the negative binomial family, the differing conclusions of the gof tests are not unexpected. This indicates that the geometric distribution may serve as a reasonable first approximation, but may still be insufficiently flexible to provide a fully satisfactory fit to the data. In addition to the conclusions based on the reported p-values, Figure 6.1 presents a graphical comparison between the observed relative frequencies and the pmf of the corresponding geometric distribution.

Dataset	$\hat{p}$	T_n	I_n	K_n	W_n	H_n	$T_n^{(2)}$	$T_n^{(3)}$	L_n
Bowel cancer	0.24	0	0	0	0	0	0	0	0
Dicentric	0.81	0.38	0.42	0.67	0.52	0.55	0.74	0.75	0.88
Primula	0.35	0.043	0.043	0.060	0.109	0.062	0.018	0.021	0.025
Earthquake	0.15	0.60	0.59	0.73	0.30	0.56	0.60	0.58	0.58

Table 6.1: Empirical p-values for testing goodness-of-fit to the geometric distribution

6.2 Applications to continuous data

The newly proposed test, along with the competitors presented in Chapter 3, is applied in this section to three real data sets involving continuous data, in order to illustrate its practical performance.

The first dataset (Table 6.2) represents the shelf life (in days) of a food product; see Gacula and Kubala, 1975. Shelf life is often assessed through failure times recorded under controlled storage conditions, but simple average-based estimates may be unreliable when these times are skewed or censored. In sensory studies, failure is commonly defined by criteria such as the just-noticeable difference (JND) or storage life, whereas procedures such as the triangle test are used to identify detectable changes in product quality. The distribution of shelf life

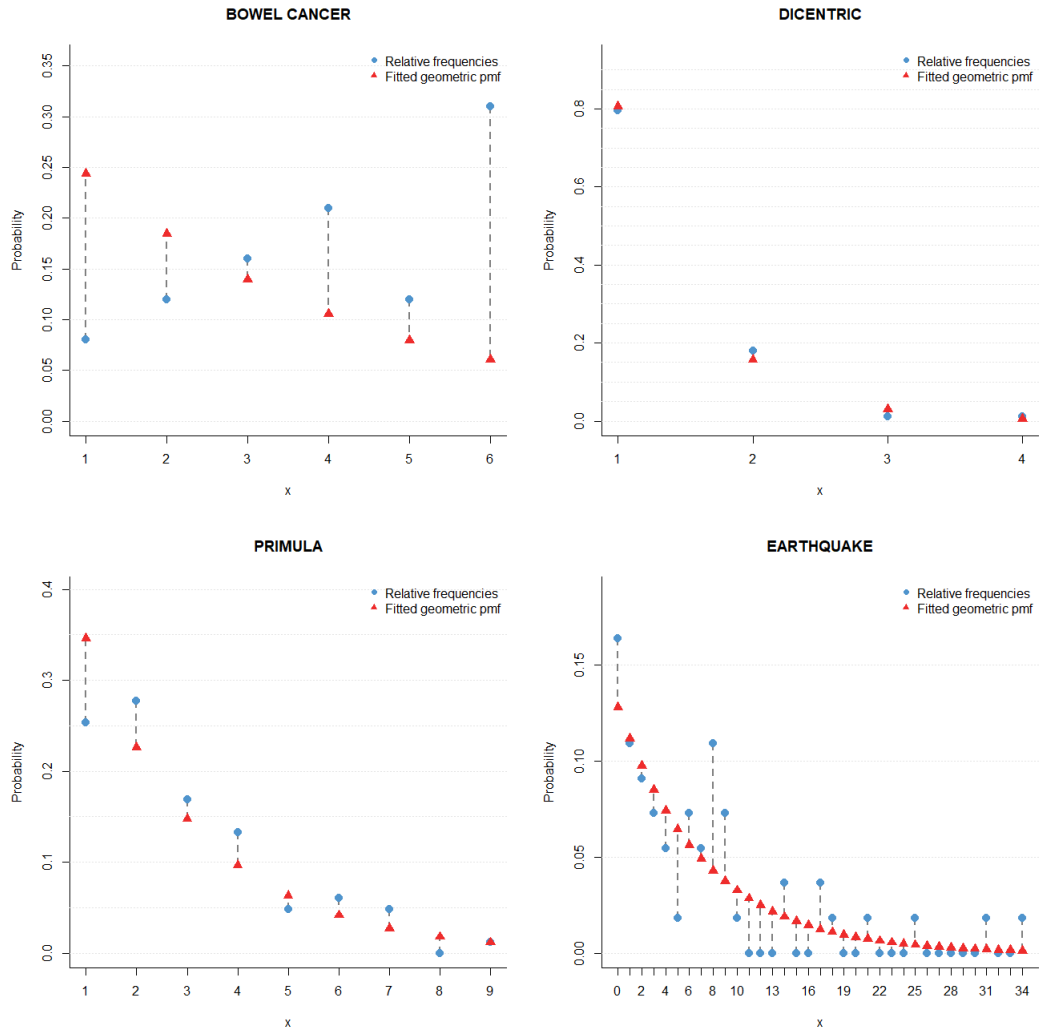


Figure 6.1: Graphical comparison of the observed relative frequencies and the fitted geometric probability mass functions

failures is therefore examined here as a basis for choosing an appropriate model for this type of data.

24, 24, 26, 26, 32, 32, 33, 33, 33, 35, 41, 42, 43,
47, 48, 48, 48, 50, 52, 54, 55, 57, 57, 57, 57, 61
Table 6.2: Dataset: *Shelf life*

The second dataset (Table 6.3) contains failure times for the right rear brakes on D9G-66A Caterpillar tractors; see Barlow and Campo, 1975. Since the brake system is a wear-sensitive service component, the recorded values represent the accumulated hours of operation before the corresponding brake unit failed. The dataset thus captures the in-service lifetime of one precisely identified part of the machine.

The final dataset (Table 6.4) refers to active repair times for an airborne transceiver (in hours); see Alven, 1964. Within the framework of maintainability analysis, active repair time denotes the portion of corrective-maintenance downtime during which repair work is actually performed, including activities such as fault verification, fault location, repair or replacement, and final testing. Thus, the observations represent the duration of the repair actions themselves, rather than the time to failure or delays due to logistical or administrative factors.

56, 83, 104, 116, 244, 305, 429, 452, 453, 503, 552, 614, 661, 673, 683,
685, 753, 763, 806, 834, 838, 862, 897, 904, 981, 1007, 1008, 1049, 1060,
1107, 1125, 1141, 1153, 1154, 1193, 1201, 1253, 1313, 1329, 1347, 1454,
1464, 1490, 1491, 1532, 1549, 1568, 1574, 1586, 1599, 1608,1723, 1769,
1795, 1927, 1957, 2005, 2010, 2016, 2022, 2037, 2065, 2096, 2139, 2150,
2156, 2160, 2190, 2210, 2220, 2248, 2285, 2325, 2337, 2351, 2437, 2454,
2546, 2565, 2584, 2624, 2675, 2701, 2755, 2877, 2879, 2922, 2986, 3092,
3160, 3185, 3191, 3439, 3617, 3685, 3756, 3826, 3995, 4007, 4159, 4300,
4487, 5074, 5579, 5623, 6869, 7739

Table 6.3: Dataset: *Tractors*

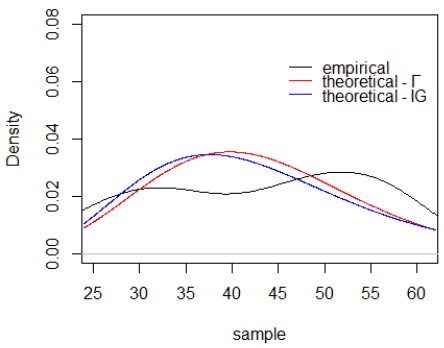
0.2, 0.3, 0.5, 0.5, 0.5, 0.5, 0.6, 0.6, 0.7, 0.7, 0.7, 0.8,
0.8, 1.0, 1.0, 1.0, 1.0, 1.1, 1.3, 1.5, 1.5, 1.5, 1.5, 2.0,
2.0, 2.2, 2.5, 2.7, 3.7, 3.0, 3.3, 3.3, 4.0, 4.0, 4.5, 4.7,
5.0, 5.4, 5.4, 7.0, 7.5, 8.8, 9.0, 10.3, 22.0, 24.5

Table 6.4: Dataset: *Transceiver*

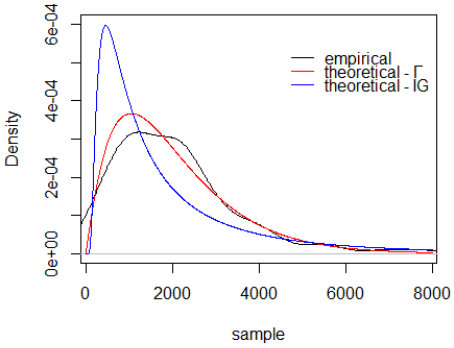
Since all these datasets represent lifetime data, the gamma and inverse Gaussian distributions are considered as null lifetime distributions for testing.

Figure 6.2 presents a kernel density estimate for each dataset, generated using the function `density` in `R` with the default kernel and bandwidth settings, together with the estimated theoretical gamma and inverse Gaussian densities. The gof tests discussed in Chapter 3 were applied to the three datasets to assess the gof for the gamma and inverse Gaussian distributions. The corresponding p-values, resulting from 10,000 bootstrap samples, are provided in Tables 6.5 and 6.6.

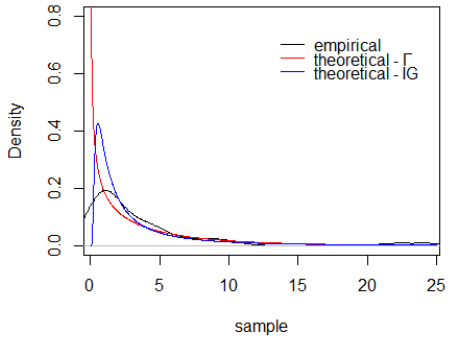
Figure 6.2 clearly shows that neither of the assumed models provides an adequate fit to the *Shelf life* dataset. However, the traditional Kolmogorov-Smirnov test does not reject the hypothesis of a gamma distribution at the 0.05 significance level. This behavior is consistent with the patterns observed in the simulation study. For this dataset, the graph suggests that the empirical density is closest



Dataset: *Shelf life*



Dataset: *Tractors*



Dataset: *Transceiver*

Figure 6.2: Density estimators for the data sets

Dataset	$\hat{\alpha}_n$	$\hat{\beta}_n$	$HBKR_n$	T_n	KS_n	CvM_n
shelf life	13.6	0.318	0.047	0.039	0.053	0.022
tractors	2.100	0.001	0.39	0.37	0.40	0.74
transceiver	0.549	0.150	$< 10^{-2}$	$< 10^{-2}$	0.01	0.05

Table 6.5: Empirical p-values for testing goodness-of-fit to the gamma distribution

Dataset	$\hat{\mu}_n$	$\hat{\lambda}_n$	$HBKR_n$	T_n	KS_n	CvM_n
shelf life	42.9	484	0.027	0.017	0.041	0.026
tractors	2024	1472	$< 10^{-2}$	$< 10^{-2}$	$< 10^{-2}$	$< 10^{-2}$
transceiver	3.62	1.66	0.70	0.63	0.90	0.83

Table 6.6: Empirical p-values for testing goodness-of-fit to the inverse Gaussian distribution

to a uniform distribution, among those considered in the simulation study, with support on $[a, b]$, where $a > 0$. When testing the null gamma hypothesis against the uniform alternative, it is evident that T_n and CvM are the most powerful tests, which corresponds with the obtained p-values. For the null inverse Gaussian hypothesis, T_n performs the best, while HBKR and CvM show similar powers, which is also reflected in the p-values.

For the *Tractors* dataset, Figure 6.2 indicates that the inverse Gaussian distribution does not provide a good fit, a conclusion supported by the test results. On the other hand, none of the tests rejects the null hypothesis of the gamma distribution, suggesting that it is a suitable alternative to the exponential distribution previously rejected in Jovanović et al., 2015.

Finally, for the *Transceiver* dataset, Figure 6.2 clearly demonstrates that the gamma distribution is not an appropriate model, a conclusion supported by all tests. The fit of the inverse Gaussian distribution is less evident from the density plot, due to the pronounced concentration of the empirical density at smaller values and the long right tail of the sample. Nevertheless, all p-values indicate that the null hypothesis should not be rejected.

6.3 Applications to circular data

This section illustrates the practical applicability of the test developed in Chapter 4, together with the corresponding competing methods, applied to several real data sets to assess the adequacy of the wrapped normal model.

The first dataset, taken from Ghosh et al., 2024, contains 70 confirmed impact craters listed in the Earth Impact Database. For each crater, detailed measurements are available to describe both the spatial and temporal characteristics of meteorite impacts on the Earth's surface. Attention is restricted here to the longitudinal component in order to investigate global patterns of crater clustering (Table 6.7). Although the Earth Impact Database is widely recognized as the principal reference source for confirmed impact structures, ongoing geological investigations may still lead to revisions of estimated ages or to the identification of additional structures.

To further illustrate the applicability of the proposed test, a second dataset from Lindecke et al., 2019 was analyzed. This study involved 26 adult *Pipistrellus pygmaeus* bats captured along the Latvian Baltic Sea coast and randomly assigned either to a control group (C), exposed to the natural sunset, or to a mirror group (M), in which the sunset azimuth was shifted by 180°. The animals were kept in

50.467	22.667	-61.700	133.133	16.900	135.200	117.583
-64.217	27.400	-115.600	-5.500	-1.417	172.000	82.017
48.033	73.450	27.500	-89.500	-81.183	111.183	-68.700
135.450	23.533	64.150	-114.000	142.833	-70.300	14.867
73.450	-64.217	-52.983	-76.017	29.667	43.717	-98.533
114.667	-109.500	-74.500	-94.550	-87.000	118.833	24.600
120.883	-63.300	-74.100	40.500	-117.633	133.583	-113.967
32.250	10.567	-74.800	-89.667	23.700	0.783	132.300
134.833	95.933	32.917	78.133	-17.067	127.783	-68.250
-10.400	129.317	-111.017	123.450	76.500	28.067	60.967

Table 6.7: Longitude of craters

their treatment cages from 30 minutes before until 30 minutes after sunset, then translocated inland, and their mean departure bearings at takeoff were recorded (Table 6.8). The gof of the resulting bearings to the wrapped normal distribution was subsequently assessed for both groups.

AdultC		
310	230	240
210	145	215
270	320	310
330	210	

(a)

AdultM		
230	140	60
60	50	40
80	110	90
270	90	210
20	90	45

(b)

Table 6.8: Takeoff orientation of Pipistrellus pygmaeus adults: control (a) vs. mirror (b) group

Another illustrative dataset, shown in Table 6.9, involves observations of 100 ants placed individually in a circular arena, where a black target served as a potential directional cue. Researchers tracked the movement paths of each ant to determine whether there was a consistent orientation toward the target. This experiment, originally conducted in Jander, 1957, provides valuable insights into the mechanisms of visual guidance and spatial orientation in ants. It serves as a foundational study in understanding insect navigation under controlled conditions.

The last dataset contains movement directions observed for tracked individuals of the sea star *Astropecten jonstoni* in the shallow coastal waters of Sardinia, in the Mediterranean Sea (Table 6.10). These data provide insight into the spatial behavior of *Astropecten jonstoni* and make it possible to assess the environmental factors that may influence its coastal orientation and migration patterns. The data

330	290	60	200	200	180	280	220	190	180
180	160	280	180	170	190	180	140	150	150
160	200	190	250	180	30	200	180	200	350
200	180	120	200	210	130	30	210	200	230
180	160	210	190	180	230	50	150	210	180
190	210	220	200	60	260	110	180	220	170
10	220	180	210	170	90	160	180	170	200
160	180	120	150	300	190	220	160	70	190
110	270	180	200	180	140	360	150	160	170
140	40	300	80	210	200	170	200	210	190

Table 6.9: Movements of ants

were collected by scuba diving, with particular attention given to the behavior of the animals after displacement. For reference, the coastal direction was taken to be approximately 0°. The dataset was reported in Upton and Fingleton, 1989 on the basis of graphical material originally presented in Pabst and Vicentini, 1978.

0	1	3	3	8	13	16	18	30	31	43
45	147	298	329	332	335	340	350	354	356	357

Table 6.10: Movements of sea stars

Table 6.11 summarizes the maximum likelihood estimates for the wrapped normal distribution, along with the associated p-values for each analyzed dataset. These results are obtained using the tests $S_{n,1}$, K_n, W_n, and $C_{n,0.5}$. The parametric bootstrap procedure is used with N = 5000 bootstrap samples. For the first

Dataset	n	$\hat{\mu}_n$	$\hat{\sigma}_n$	$S_{n,1}$	K_n	W_n	$C_{n,0.5}$
craters	70	0.27	1.63	0.02	< 0.01	< 0.01	0.01
adultC	11	-1.85	0.98	0.17	0.24	0.26	0.44
adultM	15	1.62	1.31	0.03	0.04	0.04	0.01
ants	100	3.13	1.11	< 0.01	< 0.01	< 0.01	< 0.01
sea stars	21	0.02	0.44	0.83	0.75	0.80	0.64

Table 6.11: Empirical p-values for testing goodness-of-fit to the wrapped normal distribution.

dataset, all considered tests lead to the rejection of H_0, which is in agreement with the findings of Ghosh et al., 2024, where the von Mises distribution was identified as providing an adequate fit. This agreement is not unexpected, since Table 4.3 indicates that, for the sample size and parameter estimates considered here, all tests possess sufficient power to discriminate between the wrapped normal and von Mises distributions.

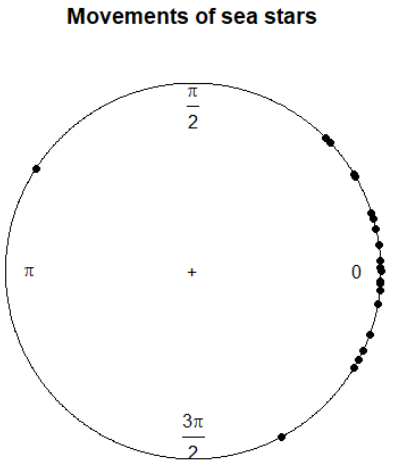


Figure 6.3: Plot of the movements of sea stars

Turning to the next dataset, none of the tests rejected the null hypothesis of a wrapped normal distribution for the adultC group. In contrast, all tests rejected the null hypothesis for the adultM group, in agreement with the original study, which reported a significant difference in orientation distributions between the two treatment groups.

For the ants dataset, all tests reject the null hypothesis, indicating that the wrapped normal distribution does not provide an adequate fit. This finding is also consistent with the conclusions reported in Ravindran, 2002, where the wrapped Cauchy distribution was identified as a more appropriate model for this dataset.

Turning to the final dataset, the corresponding p-values are $S_{n,1} = 0.2312$, $K_n = 0.0956$, $W_n = 0.0318$, and $C_{n,0.5} = 0.0056$. These results show that neither the newly proposed test nor the K_n test rejects the null hypothesis, thereby supporting the adequacy of the wrapped normal model for these data. In contrast, the remaining two tests lead to rejection of H_0. As can be seen from Figure 6.3, two observations appear to deviate noticeably from the main pattern of the sample, which may have affected the overall gof evaluation.

Accordingly, the dataset is examined by means of an outlier test for discordancy specific to wrapped normal data. The general version of this test was originally proposed by Mardia, 1975, whereas the form employed here was presented in Collett, 1980. The applied test statistic is

$$M_n = \frac{R_{(n)} - R + 1}{n - R}.$$

Here, R denotes the resultant length of the full sample, and $R_{(n)} = \max_i R_i$, where

R_i is the sample resultant length computed after omitting the ith observation θ_i. An observation is classified as an outlier when the value of M_n lies above the corresponding critical threshold, with the relevant critical values given in Rambli et al., 2012. Applying this procedure, one sea star, namely the specimen that moved markedly away from the general pattern and headed out to sea, was identified as a significant outlier (value 147°). After detecting and removing this outlier, the p-values for all tests increased significantly. These updated p-values are presented in Table 6.11.

Conclusion

This dissertation has introduced several new goodness-of-fit tests motivated by independence-type characterizations. More precisely, new tests have been developed for the geometric distribution, for a broad class of absolutely continuous distributions, and for the wrapped normal distribution in the context of circular data. In addition, a general testing approach based on linear combinations of degenerate V-statistics has been considered, providing a flexible framework for constructing new test procedures and studying their asymptotic and finite-sample properties.

First, in the Introduction, the theoretical background needed for the asymptotic analysis of the newly proposed gof tests throughout the dissertation is provided. It summarizes the main results on U- and V-statistics, their degeneracy properties, and the corresponding limiting distributions. On this basis, novel theoretical results are derived for the limiting distributions of finite linear combinations of weakly degenerate V-statistics, both in the case of known parameters and in settings involving parameter estimation. The chapter further introduces the basic concepts of U-processes and presents the characterizations of probability distributions, focusing on independence characterizations as the foundation for the gof tests developed in the dissertation.

Chapter 2 presents the gof tests proposed for the geometric distribution, together with the main conclusions arising from their theoretical and empirical analysis. In contrast to the commonly used approach of constructing statistics from the empirical counterpart of the characterizing differential equation whose solution is the pgf of the distribution under consideration, the proposed tests are based on an independence-type characterization of the geometric distribution. The resulting test statistics are based on the discrepancy between the joint V-empirical pgf and the product of the corresponding marginal V-empirical probability generating functions associated with the random variables appearing in the characterization.

The asymptotic null distributions of the proposed statistics are derived, and their almost sure limits under general conditions are established. In addition,

the consistency of parametric bootstrap estimators of the corresponding null distributions is proved. The finite-sample performance of the tests is then examined through a comprehensive empirical power study, which shows that the proposed tests behave competitively in comparison with both classical procedures and more recently developed tests.

Several directions for future research naturally arise from the presented results. One possible extension is the incorporation of weight functions, that is, the study of analogous test statistics in the Hilbert space $L^2_w[0,1]^2$, equipped with the inner product

$$\langle f, g \rangle_w = \int_0^1 \int_0^1 f(s,t) g(s,t) w(s,t)\, ds\, dt,$$

where the weight function w satisfies suitable regularity conditions. From a computational point of view, one feasible choice is

$$w(t,s) = t^a (1-t)^b s^c (1-s)^d,$$

where the positive constants a, b, c, and d may be treated as tuning parameters.

Another possible direction is to replace probability generating functions by other transformations that uniquely characterize the distribution, with appropriate modifications of the norms in the corresponding Hilbert spaces. Although such procedures may be computationally more demanding, they could potentially lead to improved power performance.

Beyond extending the applicability of the characterization considered here, it would also be of interest to investigate new characterizations analogous to those established for the geometric distribution and to develop gof tests based on them.

Chapter 3 addresses a broader characterization-based framework for gof testing. The main emphasis is placed on the construction of a general class of test statistics based on independence-type characterizations for families of absolutely continuous null distributions. Within this framework, the asymptotic properties of the proposed test are investigated, consistency against fixed alternatives is established, and finite-sample performance is assessed by means of a simulation study.

A detailed comparison with the characterization-based HBKR test and classical omnibus procedures shows that the proposed approach performs favorably across a wide range of settings. In particular, the empirical results indicate that this test often provides an improvement over traditional Kolmogorov-Smirnov and Cramér-von Mises procedures, especially for smaller sample sizes, while remaining competitive with HBKR test based on the same general principle. The

numerical study further indicates that different implementations of the same characterization can give rise to substantially different testing procedures, not only in their construction but also in their ability to detect specific alternatives.

A natural extension of the proposed methodology would be to incorporate an appropriate weight function into the test statistic. In particular, in parametric settings, such an approach would also enable a more detailed discussion of how the choice of tuning parameters affects sensitivity to specific alternatives. Further developments could also consider characterizations involving functions of dimension higher than two, which would make it possible to extend the theoretical results to a broader class of testing problems.

The next part of the dissertation, Chapter 4, concerns gof testing for circular data, with particular emphasis on the wrapped normal distribution. In this setting, a new testing procedure is developed from an independence-type characterization of the null model. The corresponding test statistic is formulated through the difference between the joint and the product of the marginal V-empirical characteristic functions associated with the characterization.

A detailed theoretical analysis is carried out for the proposed procedure. In particular, the asymptotic null distribution of the test statistic is obtained, and its almost sure behavior under general assumptions is established. Its finite-sample properties are then examined by means of an extensive simulation study, where the new test is evaluated alongside the procedure based on the L^2-type distance between empirical and theoretical characteristic functions, as well as several classical gof tests. The numerical findings show that the proposed method performs competitively across a range of alternatives.

The developments presented in this chapter also indicate several possible directions for future research. These include the study of additional characterizations analogous to those available for the wrapped normal distribution, computational refinements aimed at improving performance for larger datasets, and extensions of the proposed framework to multivariate circular data. In this way, the chapter contributes not only a new gof procedure for the wrapped normal distribution, but also a broader basis for further advances in the analysis of directional data.

Chapter 5 is devoted to gof tests constructed as linear combinations of weakly degenerate V-statistics. The asymptotic distributions of these combined statistics, derived in the introductory chapter for both known and estimated parameters, provide the theoretical basis for the methodology developed here. The chapter presents a general framework for constructing such combined tests and discusses

the choice of weights, with particular emphasis on weighting each component by the inverse of its corresponding null standard deviation.

The empirical behavior of the proposed methodology is studied through a broad simulation analysis covering a wide range of data-generating scenarios. The results show that the combined statistics provide a useful compromise when individual tests display contrasting power properties under different alternatives. At the same time, this approach provides a simple and effective alternative to tuning-parameter selection in families of test statistics indexed by such parameters. The numerical results show that combining the component statistics generally yields an improvement in power compared with using each statistic separately.

The applicability of the method is illustrated through several testing problems, including exponentiality testing, multivariate normality, and gof for selected directional distributions. The findings presented in this chapter indicate that the proposed approach is both broadly applicable and practically attractive, since it improves testing performance while avoiding computationally demanding data-driven tuning procedures.

The final chapter brings together the real-data analyses corresponding to the testing procedures developed throughout the dissertation. Its purpose is to provide a unified illustration of the practical applicability of the proposed methods and to show how they perform in concrete settings arising from different types of data.

Bibliography

Agostinelli, C. (2007). Robust estimation for circular data. *Computational Statistics & Data Analysis* 51(12), 5867–5875.

Agostinelli, C. and U. Lund (2022). *R package* `circular`*: Circular Statistics (version 0.4-95)*.

Allison, J. S. and L. Santana (2015). On a data-dependent choice of the tuning parameter appearing in certain goodness-of-fit tests. *Journal of Statistical Computation and Simulation* 85(16), 3276–3288.

Alven, W. (1964). *Reliability engineering*. New Jersey: Prentice Hall.

Anan, O., D. Böhning, and A. Maruotti (2019). On the Turing estimator in capture–recapture count data under the geometric distribution. *Metrika* 82(2), 149–172.

Arnold, B. C. (1979). Some characterizations of the Cauchy distribution. *Australian Journal of Statistics* 21(2), 166–169.

Baringhaus, L. and D. Gaigall (2015). On an independence test approach to the goodness-of-fit problem. *Journal of Multivariate Analysis* 140, 193–208.

Baringhaus, L. and N. Henze (1988). A consistent test for multivariate normality based on the empirical characteristic function. *Metrika* 35(1), 339–348.

Barlow, R. E. and R. Campo (1975). Total time on test processes and applications to failure data analysis. In: *Reliability and Fault Tree Analysis*. Philadelphia: SIAM, 451–481.

Batsidis, A., M. D. Jiménez-Gamero, and A. J. Lemonte (2020). On goodness-of-fit tests for the Bell distribution. *Metrika* 83(3), 297–319.

Billingsley, P. (1999). *Convergence of Probability Measures*. 2nd ed. New York: John Wiley & Sons.

Bliss, C. I. and R. A. Fisher (1953). Fitting the negative binomial distribution to biological data. *Biometrics* 9(2), 176–200.

Borodavka, J. I. and B. Ebner (2026). A general maximal projection approach to uniformity testing on the hypersphere. *Bernoulli* 32(2), 996 –1019.

Burgio, G. and Ya. Yu. Nikitin (2001). The combination of the sign and Wilcoxon tests for symmetry and their Pitman efficiency. In: *Asymptotic Methods in Probability and Statistics with Applications*. Boston: Birkhäuser, 395–407.

Burgio, G. and Ya. Yu. Nikitin (2003). On the combination of the sign and Maesono tests for symmetry and its efficiency. *Statistica* 63(2), 213–222.

Collett, D. (1980). Outliers in circular data. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 29(1), 50–57.

Cuparić, M., B. Milošević, and M. Obradović (2019). New L^2-type exponentiality tests. *SORT* 43(1), 25–49.

— (2022). New consistent exponentiality tests based on V-empirical Laplace transforms with comparison of efficiencies. *Revista de la Real Academia de Ciencias Exactas, Físicas y Naturales. Serie A (Matemáticas)* 116(1), 42.

D'Agostino, R. and M. Stephens, eds. (1986). *Goodness-of-fit techniques*. New York: Marcel Dekker.

Deshpande, J. V. and S. C. Kochar (1983). A linear combination of two U-statistics for testing new better than used. *Communications in Statistics - Theory and Methods* 12(2), 153–159.

Diggle, P. J., N. I. Fisher, and A. J. Lee (1985). A comparison of tests of uniformity for spherical data. *Australian Journal of Statistics* 27(1), 53 –59.

Doksum, K. and R. Thompson (1971). Power bounds and asymptotic minimax results for one-sample rank tests. *The Annals of Mathematical Statistics* 42(1), 12–34.

Dudley, R. (2002). *Real Analysis and Probability*. Cambridge: Cambridge University Press.

Ebner, B. and N. Henze (2020). Tests for multivariate normality - A critical review with emphasis on weighted L^2-statistics. *Test* 29(4), 845–892.

Ebner, B., M. D. Jiménez-Gamero, and B. Milošević (2025). Efficient eigenvalue approximation in covariance operators via Rayleigh-Ritz with statistical applications. *Statistical Papers* 66(6), 135.

Ejsmont, W., B. Milošević, and M. Obradović (2023). A test for normality and independence based on characteristic function. *Statistical Papers* 64(6), 1861–1889.

Ferguson, T. S. (1967). On characterizing distributions by properties of order statistics. *Sankhyā: The Indian Journal of Statistics, Series A (1961-2002)* 29(3), 265–278.

Fernández-de-Marcos, A. and E. García-Portugués (2023). On new omnibus tests of uniformity on the hypersphere. *Test* 32(4), 1508–1529.

Figueiredo, A. (2007). Comparison of tests of uniformity defined on the hypersphere. *Statistics & Probability Letters* 77(3), 329–334.

Fisher, R. A. (1934). *Statistical methods for research workers.* Edinburgh, London: Oliver and Boyd.

Gacula, M. and J. Kubala (1975). Statistical models for shelf life failures. *Journal of Food Science* 40(2), 404–409.

García-Portugués, E., P. L. de Micheaux, S. G. Meintanis, and T. Verdebout (2024). Nonparametric tests of independence for circular data based on trigonometric moments. *Statistica sinica* 34(2), 567–588.

García-Portugués, E. and T. Verdebout (2018). An overview of uniformity tests on the hypersphere. *arXiv preprint arXiv:1804.00286*.

Ghosh, P., D. Chatterjee, and A. Banerjee (2024). On the directional nature of celestial object's fall on the earth (Part 1: distribution of fireball shower, meteor fall, and crater on earth's surface). *Monthly Notices of the Royal Astronomical Society* 531(1), 1294–1307.

Giacomini, R., D. N. Politis, and H. White (2013). A warp-speed method for conducting Monte Carlo experiments involving bootstrap estimators. *Econometric Theory* 29(3), 567–589.

Gilitschenski, I., G. Kurz, and U. D. Hanebeck (2013). Bearings-only sensor scheduling using circular statistics. In: *Proceedings of the 16th International Conference on Information Fusion*. Istanbul: IEEE, 515–521.

Halaj, K. (2025). A goodness-of-fit test for the wrapped normal distribution based on an independence characterization. *Communications in Statistics - Simulation and Computation*, 1–17.

Halaj, K. and B. Milošević (2025). On the application of Ferguson characterization for the construction of a goodness-of-fit test for the geometric distribution. *Statistical Methods & Applications* 34(3), 545–560.

Halaj, K., B. Milošević, M. Obradović, and M. D. Jiménez-Gamero (2024). Correlation-type goodness-of-fit tests based on independence characterizations. *AStA Advances in Statistical Analysis* 108, 185–207.

Henze, N. (1996). Empirical-distribution-function goodness-of-fit tests for discrete models. *The Canadian Journal of Statistics* 24(1), 81–93.

Henze, N. and S. G. Meintanis (2005). Recent and classical tests for exponentiality: a partial review with comparisons. *Metrika* 61(1), 29–45.

Henze, N. and B. Zirkler (1990). A class of invariant consistent tests for multivariate normality. *Communications in Statistics - Theory and Methods* 19(10), 3595–3617.

Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *Annals of Mathematical Statistics*, 293–325.

Hoeffding, W. (1961). *The strong law of large numbers for U-statistics*. Tech. rep. 302. Department of Statistics, North Carolina State University.

Jammalamadaka, S. R., M. D. Jiménez-Gamero, and S. G. Meintanis (2019). A class of goodness-of-fit tests for circular distributions based on trigonometric moments. *SORT* 43(2), 317–336.

Jammalamadaka, S. R. and A. SenGupta (2001). *Topics in Circular Statistics*. Singapore: World Scientific.

Jander, R. (1957). Die optische Richtungsorientierung der Roten Waldameise (Formica Rufa L.) *Journal of Comparative Physiology* 40, 162–238.

Janssen, A. (2000). Global power functions of goodness of fit tests. *The Annals of Statistics* 28(1), 239–253.

Jiménez-Gamero, M. D., B. Milošević, and M. Obradović (2020). Exponentiality tests based on Basu characterization. *Statistics* 54(4), 714–736.

Jiménez-Gamero, M. D. and M. V. Alba-Fernández (2021). A test for the geometric distribution based on linear regression of order statistics. *Mathematics and Computers in Simulation* 186, 103–123.

Jiménez-Gamero, M. D., M. V. Alba-Fernández, J. Muñoz-García, and Y. Chalco-Cano (2009). Goodness-of-fit tests based on empirical characteristic functions. *Computational Statistics & Data Analysis* 53(12), 3957–3971.

Jona Lasinio, G., A. Gelfand, and M. Jona Lasinio (2012). Spatial analysis of wave direction data using wrapped Gaussian processes. *The Annals of Applied Statistics* 6(4), 1478–1498.

Jovanović, M., B. Milošević, Ya. Yu. Nikitin, M. Obradović, and K. Yu. Volkova (2015). Tests of exponentiality based on Arnold-Villasenor characterization and their efficiencies. *Computational Statistics & Data Analysis* 90, 100–113.

Kagan, Y. Y. (2010). Statistical distributions of earthquake numbers: consequence of branching process. *Geophysical Journal International* 180(3), 1313–1328.

Khatri, C. G. (1962). A characterization of the inverse Gaussian distribution. *Annals of Mathematical statistics* 33(2), 800–803.

Klar, B. (1999). Goodness-of-fit tests for discrete models based on the integrated distribution function. *Metrika* 49, 53–69.

Korolyuk, V. S. and Y. V. Borovskikh (1994). *Theory of U-statistics*. Dordrecht: Kluwer Academic Publishers.

Korolyuk, V. S. and Y. V. Borovskikh (2013). *Theory of U-statistics*. Vol. 273. Springer Science & Business Media.

Kuiper, N. H. (1960). Tests concerning random points on a circle. In: *Proceedings of the Koninklijke Nederlandse Akademie van Wetenschappen, Series A*. Vol. 63. (1). Amsterdam: North-Holland Publishing Company, 38–47.

Kundu, S., S. Majumdar, and K. Mukherjee (2000). Central limit theorems revisited. *Statistics & Probability Letters* 47(3), 265–275.

Kurz, G., F. Faion, and U. D. Hanebeck (2013). Constrained object tracking on compact one-dimensional manifolds based on directional statistics. In: *International Conference on Indoor Positioning and Indoor Navigation*. Montbéliard, France: IEEE, 1–9.

Lévy, P. (1939). L'addition des variables aléatoires définies sur une circonférence. *Bulletin de la Société mathématique de France* 67, 1–41.

Lindecke, O., A. Elksne, R. A. Holland, G. Pētersons, and C. C. Voigt (2019). Experienced migratory bats integrate the sun's position at dusk for navigation at night. *Current Biology* 29(8), 1369–1373.

Loughin, T. M. (2004). A systematic comparison of methods for combining p-values from independent tests. *Computational Statistics & Data Analysis* 47(3), 467–485.

Lukacs, E. (1955). A characterization of the gamma distribution. *The Annals of Mathematical Statistics* 26(2), 319–324.

Manzotti, A. and A. J. Quiroz (2001). Spherical harmonics in quadratic forms for testing multivariate normality. *Test* 10, 87–104.

Mardia, K. V. and P. E. Jupp (2000). *Directional Statistics*. New York: John Wiley & Sons.

Mardia, K. V. (1975). Statistics of directional data. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)* 37(3), 349–371.

McCullagh, P. (1996). Möbius transformation and Cauchy parameter estimation. *The Annals of Statistics* 24(2), 787–808.

Meintanis, S. G., B. Milošević, M. Obradović, and M. Veljović (2024a). Goodness-of-fit tests for the multivariate Student-t distribution based on iid data, and for GARCH observations. *Journal of Time Series Analysis* 45(2), 298–319.

Meintanis, S. G., Ya. Yu. Nikitin, and A. V. Tchirina (2007). Tests for exponentiality against NBRUE alternative life distributions. *International Journal of Statistics and Management Systems* 2(1-2), 31–43.

Meintanis, S. G. (2004). Goodness-of-fit tests for the logistic distribution based on empirical transforms. *Sankhyā: The Indian Journal of Statistics (2003-2007)* 66(2), 306–326.

Meintanis, S. G., B. Milošević, and M. D. Jiménez-Gamero (2024b). Goodness-of-fit tests based on the min-characteristic function. *Computational Statistics & Data Analysis* 197, 107988.

Meintanis, S. G., J. Ngatchou-Wandji, and E. Taufer (2015). Goodness-of-fit tests for multivariate stable distributions based on the empirical characteristic function. *Journal of Multivariate Analysis* 140, 171–192.

Milošević, B. (2017). Some recent characterization based goodness of fit tests. In: *Proceedings of the 20th European Young Statisticians Meeting*. Uppsala: Uppsala University, 67–73.

Milošević, B. and M. Obradović (2016). Two-dimensional Kolmogorov-type goodness-of-fit tests based on characterisations and their asymptotic efficiencies. *Journal of Nonparametric Statistics* 28(2), 413–427.

Milošević, B., M. D. Jiménez-Gamero, and M. V. Alba-Fernández (2021). Quantifying the ratio-plot for the geometric distribution. *Journal of Statistical Computation and Simulation* 91, 2153–2177.

Mudholkar, G. S., R. Natarajan, and Y. P. Chaubey (2001). A goodness-of-fit test for the inverse Gaussian distribution using its independence characterization. *Sankhyā: The Indian Journal of Statistics, Series B (1960-2002)* 63(3), 362–374.

Ndwandwe, L. M., J. S. Allison, M. Smuts, and J. Visagie (2023). On a new class of tests for the Pareto distribution using Fourier methods. *Stat* 12(1), e566.

Novoa-Muñoz, F. and M. D. Jiménez-Gamero (2014). Testing for the bivariate Poisson distribution. *Metrika* 77, 771–793.

Pabst, B. and H. Vicentini (1978). Dislocation experiments in the migrating sea star Astropecten jonstoni. *Marine Biology* 48, 271–278.

Puri, P. S. and H. Rubin (1970). A characterization based on the absolute difference of two iid random variables. *The Annals of Mathematical Statistics* 41(6), 2113–2122.

R Core Team (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria.

Rambli, A., S. Ibrahim, M. I. Abdullah, A. G. Hussin, and I. Mohamed (2012). On discordance test for the wrapped normal data. *Sains Malaysiana* 41(6), 769–778.

Ravindran, P. (2002). *Bayesian Analysis of Circular Data Using Wrapped Distributions*. PhD thesis. Faculty of North Carolina State University.

Rizzo, M. L. (2002). *A new rotation invariant goodness-of-fit test*. PhD thesis. Bowling Green State University.

Rueda, R. and F. O'Reilly (1999). Tests of fit for discrete distributions based on the probability generating function. *Communications in Statistics - Simulation and Computation* 28(1), 259–274.

Rukhin, A. L. (1972). Some statistical decisions about distributions on a circle for large samples. *Sankhyā: The Indian Journal of Statistics, Series A (1961-2002)* 34(3), 243–250.

Schneemeier, W. (1989). Weak convergence and Glivenko-Cantelli results for empirical processes of U-statistic structure. *Stochastic Processes and their Applications* 33(2), 325–334.

Serfling, R. (1980). *Approximation theorems of mathematical statistics*. New York: John Wiley & Sons.

Sexton, J., R. Blomhoff, A. Karlsen, and P. Laake (2012). Adaptive combination of dependent tests. *Computational Statistics & Data Analysis* 56(6), 1935–1943.

St. John, P. C., S. R. Taylor, J. H. Abel, and F. J. Doyle (2014). Amplitude metrics for cellular circadian bioluminescence reporters. *Biophysical Journal* 107(11), 2712–2722.

Stute, W., W. González Manteiga, and M. Presedo Quindimil (1993). Bootstrap based goodness-of-fit tests. *Metrika* 40(3-4), 243–256.

Tenreiro, C. (2009). On the choice of the smoothing parameter for the BHEP goodness-of-fit test. *Computational Statistics & Data Analysis* 53(4), 1038–1053.

— (2017). A new test for multivariate normality by combining extreme and nonextreme BHEP tests. *Communications in Statistics - Simulation and Computation* 46(3), 1746–1759.

— (2019). On the automatic selection of the tuning parameter appearing in certain families of goodness-of-fit tests. *Journal of Statistical Computation and Simulation* 89, 1780–1797.

Torabi, H., N. H. Montazeri, and A. Grané (2018). A wide review on exponentiality tests and two competitive proposals with application on reliability. *Journal of Statistical Computation and Simulation* 88(1), 108–139.

Upton, G. J. and B. Fingleton (1989). *Spatial data analysis by example. Volume 2. Categorical and directional data*. New York: John Wiley.

van der Vaart, A. (1998). *Asymptotic statistics*. Cambridge: Cambridge University Press.

van der Vaart, A. and J. Wellner (1996). *Weak Convergence and Empirical Processes*. New York: Springer.

Van Zwet, W. and J Oosterhoff (1967). On the combination of independent test statistics. *The Annals of Mathematical Statistics* 38(3), 659–680.

Watson, G. S. (1961). Goodness-of-fit tests on a circle. *Biometrika* 48(1/2), 109–114.

Wilson, P. and J. Einbeck (2023). A consistent way to define p-values. In: *Proceedings of the 37th IWSM*. Dortmund: TU Dortmund University, 658–663.

Wintner, A. (1947). On the shape of the angular case of Cauchy's distribution curves. *The Annals of Mathematical Statistics* 18(4), 589–593.

Yamato, H., M. Kondo, and K. Toda (2005). Asymptotic properties of linear combinations of U-statistics with degenerate kernels. *Journal of nonparametric statistics* 17(2), 187–199.

Yu, K., Q. Li, A. W. Bergen, R. M. Pfeiffer, P. S. Rosenberg, N. Caporaso, P. Kraft, and N. Chatterjee (2009). Pathway analysis by adaptive combination of P-values. *Genetic Epidemiology* 33(8), 700–709.

Biography

Katarina Halaj Mileusnić was born in Belgrade on November 17, 1993. She completed elementary school "Desanka Maksimović" in Zemun, where she was awarded the Vuk Karadžić diploma. She was a member of the Center for Talented Youth and participated in municipal, regional and national competitions, occasionally achieving award-winning results. She continued her education at Ninth Gymnasium "Mihailo Petrović Alas" in New Belgrade.

In 2012, she enrolled at the Faculty of Mathematics, University of Belgrade, where she completed Bachelor's studies in Statistics, Actuarial and Financial Mathematics in 2016 with a GPA of 9.21/10.00. She continued her studies there, earning a Master's degree in 2017 with a GPA of 10.00/10.00. Her master's thesis, entitled "Portfolio optimization based on value at risk", was written under the supervision of Professor Bojana Milošević, PhD. Later that year, she enrolled in the PhD program in Mathematics at the same faculty and subsequently passed all required exams with a GPA of 10.00/10.00.

Since 2017, she has been employed at the Faculty of Transport and Traffic Engineering, University of Belgrade, first as a Teaching Associate and, since 2018, as a Teaching Assistant. During the period 2022–2023, she was also engaged as a Teaching Assistant at the Faculty of Pharmacy, University of Belgrade.

She has published four papers in SCI-indexed journals, while two additional papers are currently under review. Her research has been presented at several international conferences, including an invited talk at the European Young Statisticians Meeting in Ljubljana in 2023. In 2025, she served as a member of the organizing committee of the 24th European Young Statisticians Meeting, held in Torino, Italy.

Her academic activities also include participation in international training courses in statistics and data science, as well as involvement in national and bilateral research projects. The main areas of her research include nonparametric statistics, goodness-of-fit testing, independence characterizations, and circular and directional data analysis.

Прилог 1.

Изјава о ауторству

Потписани-а Катарина Халај Милеуснић

број уписа 2009/2017

Изјављујем

да је докторска дисертација под насловом

Novel goodness-of-fit tests based on independence-type characterizations
(Нови тестови сагласности засновани на карактеризацијама независности статистика)

- резултат сопственог истраживачког рада,
- да предложена дисертација у целини ни у деловима није била предложена за добијање било које дипломе према студијским програмима других високошколских установа,
- да су резултати коректно наведени и
- да нисам кршио/ла ауторска права и користио интелектуалну својину других лица.

Потпис докторанда

У Београду, 13.8.2026.

К Халај Милеуснић

Прилог 2.

Изјава о истоветности штампане и електронске верзије докторског рада

Име и презиме аутора Катарина Халај Милеуснић

Број уписа 2009/2017

Студијски програм Математика

Наслов рада Novel goodness-of-fit tests based on independence-type characterizations (Нови тестови сагласности засновани на карактеризацијама независности статистика)

Ментор проф. др Бојана Милошевић

Потписани Катарина Халај Милеуснић

изјављујем да је штампана верзија мог докторског рада истоветна електронској верзији коју сам предао/ла за објављивање на порталу **Дигиталног репозиторијума Универзитета у Београду.**

Дозвољавам да се објаве моји лични подаци везани за добијање академског звања доктора наука, као што су име и презиме, година и место рођења и датум одбране рада.

Ови лични подаци могу се објавити на мрежним страницама дигиталне библиотеке, у електронском каталогу и у публикацијама Универзитета у Београду.

Потпис докторанда

У Београду, 13.8.2026.

К Халај Милеуснић

Прилог 3.

Изјава о коришћењу

Овлашћујем Универзитетску библиотеку „Светозар Марковић“ да у Дигитални репозиторијум Универзитета у Београду унесе моју докторску дисертацију под насловом:

Novel goodness-of-fit tests based on independence-type characterizations
(Нови тестови сагласности засновани на карактеризацијама независности статистика)

која је моје ауторско дело.

Дисертацију са свим прилозима предао/ла сам у електронском формату погодном за трајно архивирање.

Моју докторску дисертацију похрањену у Дигитални репозиторијум Универзитета у Београду могу да користе сви који поштују одредбе садржане у одабраном типу лиценце Креативне заједнице (Creative Commons) за коју сам се одлучио/ла.

1. Ауторство

2. Ауторство - некомерцијално

3. Ауторство – некомерцијално – без прераде

4. Ауторство – некомерцијално – делити под истим условима

5. Ауторство – без прераде

6. Ауторство – делити под истим условима

(Молимо да заокружите само једну од шест понуђених лиценци, кратак опис лиценци дат је на полеђини листа).

Потпис докторанда

У Београду, 13.8.2026.

К.Халај Милеуснић

1. Ауторство - Дозвољавате умножавање, дистрибуцију и јавно саопштавање дела, и прераде, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце, чак и у комерцијалне сврхе. Ово је најслободнија од свих лиценци.

2. Ауторство – некомерцијално. Дозвољавате умножавање, дистрибуцију и јавно саопштавање дела, и прераде, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце. Ова лиценца не дозвољава комерцијалну употребу дела.

3. Ауторство - некомерцијално – без прераде. Дозвољавате умножавање, дистрибуцију и јавно саопштавање дела, без промена, преобликовања или употребе дела у свом делу, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце. Ова лиценца не дозвољава комерцијалну употребу дела. У односу на све остале лиценце, овом лиценцом се ограничава највећи обим права коришћења дела.

4. Ауторство - некомерцијално – делити под истим условима. Дозвољавате умножавање, дистрибуцију и јавно саопштавање дела, и прераде, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце и ако се прерада дистрибуира под истом или сличном лиценцом. Ова лиценца не дозвољава комерцијалну употребу дела и прерада.

5. Ауторство – без прераде. Дозвољавате умножавање, дистрибуцију и јавно саопштавање дела, без промена, преобликовања или употребе дела у свом делу, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце. Ова лиценца дозвољава комерцијалну употребу дела.

6. Ауторство - делити под истим условима. Дозвољавате умножавање, дистрибуцију и јавно саопштавање дела, и прераде, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце и ако се прерада дистрибуира под истом или сличном лиценцом. Ова лиценца дозвољава комерцијалну употребу дела и прерада. Слична је софтверским лиценцама, односно лиценцама отвореног кода.