On the Goodness-of-fit Testing Based on the Kolmogorov-Smirnov and the Hausdorff Distances

Yun Jia

Submitted to City St George's, University of London in fulfillment
of the requirements for the award of the degree of Doctor of Philosophy.

Faculty of Actuarial Science and Insurance
Bayes Business School, City St George's, University of London

Primary Supervisor:	**VLADIMIR K. KAISHEV**
Secondary Supervisor:	**DIMITRINA S. DIMITROVA**
Examining Committee:	**GIOVANNI URGA**
	TENGYAO WANG

Thesis Submission Date: September 2025

I, Yun Jia, declare that I have produced this thesis without the assistance of any third parties other than the coauthors of the papers. Except where specific reference is made to the work of others, the contents of this thesis are original and have not previously been submitted in identical or similar form to any other degree or qualification in this, or any other university. This thesis work was conducted from October 2020 to June 2025 under the supervision of Dr. Dimitrina S. Dimitrova and Prof. Vladimir K. Kaishev at Bayes Business School, City St George's, University of London.

Yun Jia

September 2025

Contents

List of Figures

List of Tables

Acknowledgements

This PhD research project has been funded by way of a bursary from the Faculty of Actuarial Science and Insurance at Bayes Business School, City St George's, University of London.

First and foremost, I would like to express my deepest gratitude to my supervisors, Dr. Dimitrina Dimitrova and Professor Vladimir Kaishev. Their guidance and mentorship have been invaluable not only throughout my PhD, but also in shaping my approach to life. Their support refined my academic writing and deepened my understanding of research in statistics and actuarial science. They have been my steadfast anchors, enabling me to complete this honorable journey.

I would also like to acknowledge with utmost respect and gratitude Professor Zvetan G. Ignatov (1942–2024), who sadly passed away before this work was completed. His profound knowledge of probability and statistics helped me deepen my understanding of this field of mathematics.

I am sincerely grateful to Professor Yusen Kwoh, whose inspiring mentorship during my bachelor's studies sparked my interest in economic research. Without his guidance and encouragement, I would not have started my PhD journey.

My thanks also go to Pengcheng Ma, Zheng Zhong, my other friends and my family, whose unwavering support and companionship have brought me strength and joy. I owe particular thanks to my parents for their unconditional love and for instilling in me the values that have made me a better person.

My dearest appreciation goes to my beloved, Jiaxing Li. In the probability space of my life, she is the condition that ensures all probability measures are tight — keeping every moment bounded within a compact set of warmth and happiness.

Abstract

In this thesis, we focus on the goodness-of-fit (GoF) testing problem. Although there are a number of classical GoF tests proposed in the literature, there is no 'best' test that suits all purposes and possesses all the desirable properties. In this thesis, we investigate in detail the properties of two of these classical tests, namely the Kolmogorov-Smirnov and the Kuiper tests, and provide efficient and exact numerical methods to compute their p-values. As known, the latter tests are ordinal and that affects their power especially in the tails. Furthermore, we propose a new H test based on the Hausdorff distance that depends on both the ordinate and abscissa coordinates. As a result of that and of the fact that it is location invariant but scale dependent, we are able to show that its power can be optimized by appropriately selecting the scale coefficient. We illustrate the enhanced power of the H test in numerous numerical examples both in the one-sample univariate and in the two-sample multivariate settings. More precisely, we show that the H test outperforms classical alternatives like Kolmogorov-Smirnov (KS), Cramer-von Mises (CvM) and Anderson-Darling (AD) in terms of power in the univariate case, and also the Ball Divergence, Maximum Mean Discrepancy, Cross Match, the Nearest Neighbor, and some other tests in the bivariate case. Last but not least, we investigate the theoretical properties of the H test and its p-values both for finite samples and asymptotically. We further provide useful results that allow the numerical evaluation of the H test, its p-values, exact Bahadur slope and asymptotic power.

Chapter 1

Introduction

This thesis focuses on the goodness-of-fit testing problems. Goodness-of-fit testing is about checking whether a pre-specified hypothetical distribution fits a random sample, or whether two random samples come from a common but unspecified distribution. It is a fundamental statistical task underpinning model validation and data comparisons that are widely used in any area of research and project development where big volumes (samples) of data on one or more variables of interest are collected and analyzed.

There are many goodness-of-fit test statistics in the literature among which the classical Kolmogorov-Smirnov (KS), the Kuiper, the Cramer von Mises (CvM), the Anderson-Darling (AD) and the Wasserstein (W) test more recently considered in del Barrio et al. (1999) and del Barrio et al. (2000). The latter tests have gained great popularity and have been widely applied in almost any field where data is collected and analysed such as, astronomy (McQuillan et al., 2013), social sciences (Salman et al., 2015), pattern recognition (Alzubaidi and Kalita, 2016), machine learning (Gretton et al., 2012) etc., to name only a few. More recently, researchers have started to consider the goodness-of-fit test statistic in multi-dimensional, including the run tests based on the minimal spanning tree due to Friedman and Rafsky (1979), and on the shortest Hamiltonian path proposed by Biswas et al. (2014), the Wasserstein test (Hundrieser et al., 2024), the Ball divergence test (Pan et al., 2018), the Maximum Mean Discrepancy test (Gretton et al., 2012), the Cross Match test (Rosenbaum, 2005) and the Schilling-Henze Nearest Neighbor test (c.f. Schilling, 1986; Henze, 1988).

Although many tests are being proposed, there is no universally recognized

'best' tests that suits all purposes and possesses all the desirable properties. The use of a particular statistic is usually a trade-off between power and computational feasibility. For example, the widely used KS test, based on the supremum distance, is readily understood graphically, is easy to evaluate, and is distribution-free when the null is continuous. However, it is less sensitive in the tails and generally has lower power (see e.g. Mason and Schuenemeyer, 1983; Feigelson and Babu, 2020), making it less efficient for tail comparisons, which are particularly important in extreme value applications. The AD and CvM tests, based on the L^2-distance, are also distribution-free and have high power in many cases but have not been introduced for discrete or mixed null distributions. The W test, based on the L^2-Wasserstein distance, also has high power in many settings but is not distribution-free and becomes computationally demanding, especially in higher dimensions where its evaluation can be heavy without entropic regularization (see Cuturi, 2013). Graph-based multivariate tests, such as run statistics, face similar challenges, as their evaluation can be non-deterministic polynomial time (NP) problems.

Moreover, as noted by Janssen (2000), no single test can pay equal attention to an infinite number of orthogonal alternatives, meaning there is no universally powerful test.

The lack of universally powerful tests and the limitations of existing goodness-of-fit tests motivatives further research. In this thesis, we investigate the properties of two of the existing tests, i.e. the Kolmogorov-Smirnov and the Kuiper tests and propose a new test that is based on the Hausdorff distance.

In Chapter 2, we show that the difference between the critical values of the permutation KS test and the KS test defined in terms of the ecdfs of the two samples is asymptotically negligible which therefore applies to the difference between their p-values. We use this result to develop a numerically efficient recurrence method for computing p-values of the two sample KS test. A similar method but for the unweighted KS test has been independently considered by Nikiforov (1994), Hilton et al. (1994), and Schröer and Trenkler (1995), but without any theoretical justification. We have completed and generalized the work of these authors as follows: 1) we provide the missing theoretical justification of why such a recurrence approach is valid; 2) we allow flexible choices of the weight function of the KS test, leading

to better power; 3) we extend the method to compute p-values of the two sample Kuiper test, thus covering the case of data with ties; 4) we show how the recurrence can be adjusted in the spirit of Viehmann (2021), leading to better accuracy. We also derive a closed-form expression for the asymptotic distribution of the two-sample KS test, assuming jumps in the null distribution and illustrate its efficiency in computing p-values.We also give an extensive overview of the existing statistical software, computing KS and Kuiper p-values and compare it with the performance of our proposed methods, implemented in the R functions KS2sample and Kuiper2sample. The latter can be useful for researchers from various fields, who analyze data, and perform distributional hypothesis testing, applying KS and Kuiper p-values.

In Chapter 3, we consider the Hausdorff metric and its use in measuring the distance between an empirical and a theoretical cumulative distribution function (cdf). We propose a corresponding one-sample Hausdorff goodness-of-fit test statistic, the H test, give its geometric interpretation, and a method to evaluate it. We show that its exact and asymptotic p-values can be expressed correspondingly as rectangle probability and as double boundary crossing probability with respect to a Brownian bridge. Efficient numerical methods for computing the p-values, the exact Bahadur slope, and asymptotic power of the H test are also provided.

We also show that the p-values of the H test are not invariant under scale transformation. Investigating theoretically this scale dependence, we find that by appropriately selecting the scale coefficient, the power of the H test can be controlled and optimized. This is an important feature which other tests such as Kolmogorov-Smirnov (KS), Cramer von Mises (CvM) and the Anderson-Darling (AD) do not possess. In particular, based on synthetic and real data examples, we demonstrate that when testing goodness-of-fit in the tail, the power and tail sensitivity of the scale-tuned H test is higher than the power of the KS, CvM and AD tests. All these properties make the H test a competitive alternative to existing goodness-of-fit tests.

As an extension of Chapter 3, in Chapter 4, we explore the use of the Hausdorff metric between possibly multivariate empirical cumulative distribution functions with the purpose of testing for goodness-of-fit. As in the one-sample case, the clas-

sical tests such as KS, CvM, and AD have appealing properties in the univariate setting but lose efficiency for tail differences. Existing multivariate tests, including Wasserstein- and run-based methods, can achieve high power but are computationally demanding. Since no hypothesized null distribution is available, extending the existing approach in Chapter 3 to the two-sample multivariate setting is challenging. We address these challenges as follows. We introduce an explicit and computable representation of the two-sample Hausdorff (H) statistic with a geometric interpretation as the edge of the largest hypercube that can be inscribed between the two ecdfs. We propose a permutation version of H and establish its asymptotic equivalences in terms of power and type I error, under the null and (fixed or contiguous) alternative. Based on this, we develop a method to compute the exact and asymptotic p-values of the H statistic. In view of the scale dependence of H, we propose a rule for selecting the scale coefficient, so as to optimize its power. Last but not least, we give some useful properties of H including its Lipschitz continuity, qualitative robustness and connections to the Lévy-Prokhorov metric and the KS test. We demonstrate based on numerical examples that the scale-tuned Hausdorff test outperforms the major competitors in terms of power in the univariate and bivariate cases.

1.1 Publications arising from this thesis

Chapter 2: On the Efficient Exact Calculation of p-values of the Two-sample Kolmogorov-Smirnov and Kuiper Tests.

This chapter is based on the paper:

Dimitrina S. Dimitrova, Yun Jia and Vladimir K. Kaishev (2025). Efficient Exact Calculation of p-values of the Two-sample Kolmogorov-Smirnov and Kuiper Tests. *submitted*.

Chapter 3: On a New One-sample Goodness-of-fit Test based on the Hausdorff Metric.

This chapter is based on the paper:

Dimitrina S. Dimitrova, Yun Jia and Vladimir K. Kaishev (2025). On a One Sample Goodness-of-Fit Test Based on the Hausdorff Metric. *submitted*.

Chapter 4: On a New Two-sample Multivariate Goodness-of-fit Test based on the Hausdorff Metric.

This chapter is based on the paper:

Dimitrina S. Dimitrova, Yun Jia and Vladimir K. Kaishev (2025). On a Two-sample Multivariate Goodness-of-fit Test based on the Hausdorff Metric. *near submission.*

Chapter 2

On the Efficient Exact Calculation of p-values of the Two-sample Kolmogorov-Smirnov and Kuiper Tests

This chapter is based on the paper:

Dimitrina S. Dimitrova, Yun Jia and Vladimir K. Kaishev (2025). Efficient Exact Calculation of p-values of the Two-sample Kolmogorov-Smirnov and Kuiper Tests. *submitted*.

Abstract

We show that the difference between the critical values of the permutation KS test and the KS test defined in terms of the ecdfs of the two samples is asymptotically negligible which therefore applies to the difference between their p-values. We use this result to develop a numerically efficient recurrence method for computing p-values of the two sample KS test. A similar method but for the unweighted KS test has been independently considered by Nikiforov (1994), Hilton et al. (1994), and Schröer and Trenkler (1995), but without any theoretical justification.

We have completed and generalized the work of these authors as follows: 1) we provide the missing theoretical justification of why such a recurrence approach is valid; 2) we allow flexible choices of the weight function of the KS test, leading to better power; 3) we extend the method to compute p-values of the two sample Kuiper test, thus covering the case of data with ties; 4) we show how the recurrence can be adjusted in the spirit of Viehmann (2021), leading to better accuracy.

We also derive a closed-form expression for the asymptotic distribution of the

two-sample KS test, assuming jumps in the null distribution and illustrate its efficiency in computing p-values.

We finally give an extensive overview of the existing statistical software, computing KS and Kuiper p-values and compare it with the performance of our proposed methods, implemented in the R functions `KS2sample` and `Kuiper2sample`. The latter can be useful for researchers from various fields, who analyze data, and perform distributional hypothesis testing, applying KS and Kuiper p-values.

2.1 Introduction

In almost any area of research and project development such as, economics, natural sciences, industry, engineering, insurance, banking and finance, big volumes (samples) of data on one or more variables of interest are collected and analyzed. Such tasks fall within the rapidly growing field of data science (also big data statistics), which underpins the development and application of machine learning methods (see e.g., Gretton et al., 2012). In all these areas, it is often necessary to test whether the data samples, usually analyzed by pairs, come from an (unspecified) common probability distribution. There are various goodness-of-fit tests (see e.g. Meintanis et al., 2024; Borrajo et al., 2024) that are used for the purpose, among which two widely used are the one and two sample Kolmogorov-Smirnov (KS) tests (see e.g. Dimitrova et al., 2020; Moscovich, 2023; Nikiforov, 1994; Viehmann, 2021), and the Kuiper test (Paltani, 2004).

The two-sample KS test is defined as the maximum of the absolute value of the difference between the empirical distribution functions of the two samples. Since this maximum tends to be achieved around the center of the distribution, the KS test is more sensitive there, which is considered as one of its limitations. At the same time the KS-test is simple and intuitive and is therefore one of the most popular goodness-of-fit tests applied in numerous fields, e.g., in astronomy, (see Feigelson and Babu, 2020), internet behaviour analysis, (Mousavi et al., 2022), physics, (Arsioli and Dedin, 2020), hydrology, (Zhang et al., 2022), to name only a few.

The two sample Kuiper test represents an extension of the KS test that is equally sensitive both in the center and the tails of the underlying distribution. This stems from its definition as the sum of the absolute values of the most positive and most negative differences between the empirical distribution functions of the

two samples. In addition, the Kuiper test is also better suited for goodness-of-fit testing with seasonal and circular data. As a consequence, the Kuiper test has also gained significant popularity and is being widely used, as evidenced by the numerous references in the literature, see e.g., applications in finance, (Yung et al., 2008), genetics, (Dudbridge, 2006), neural-networks, (Mouli et al., 2019), astronomy (Lin et al., 2022), brain-machine interface (Chhatbar and Francis, 2013), the geology of earthquakes (Kossobokov and Panza, 2020), and many others.

All these examples represent just a small selection from the very large volume of applications, confirming the importance of providing efficient means for computing the two-sample KS and Kuiper tests and their p-values. The p-value is defined as the probability that the (KS or Kuiper) test statistic is greater than or equal to a specified fixed value.

Our three major goals in this chapter are: 1) to present efficient numerical methods for computing exact p-values of the two-sample (one-sided or two-sided) KS test and the two-sample Kuiper test, when sample sizes are finite; 2) to give a closed form formula for the asymptotic distribution of the two sample KS test, as the two sample sizes go to infinity, for the case of ties in the observations; 3) to investigate the numerical performance of both asymptotic and finite sample size methods and compare the latter with other existing statistical software. For the purpose we implement our finite sample size methods in the R functions, `KS2sample` and `Kuiper2sample`, and give an extensive overview of the existing implementations, which we believe will be helpful for the practitioners using these goodness-of-fit tests.

Computing acurate p-values is crucial in testing the null hypothesis that samples have a common (but unknown) underlying distribution. When the underlying distributions are continuous, Hodges (1958), Kim (1969) and Kim and Jennrich (1973) have given recurrence formulas to compute the exact unweighted KS p-values. Maag and Stephens (1968) and Hirakawa (1973) give formulas to compute the exact Kuiper p-values when sample sizes are equal. Their formulas are valid, since both the KS and Kuiper tests are distribution-free for continuous underlying distributions. However, in the case when the samples come from two arbitrary distributions, computing their p-values are not directly possible since the distribution of the two-sample KS/Kuiper test is unknown. In order to overcome this difficulty, we consider an alternative for-

mulation of the KS/Kuiper test, referred to as the permutation KS/Kuiper test. The latter is defined over the space of all possible pairs of samples, randomly drawn without replacement from the pooled sample.

We should point out that when the underlying distributions are non-continuous, Nikiforov (1994), Hilton et al. (1994), and Schröer and Trenkler (1995) have independently given similar formulas to compute the KS p-values conditioned on the sample realization. However, Hilton et al. (1994) has noted the computational difficulty in recovering the unconditional p-values from the conditional p-values. Schröer and Trenkler (1995) treated the conditional p-values as unconditional and referred vaguely to the "principle of randomization tests". While it is valid in view of Theorem 3.2 of Hoeffding (1952) when the null hypothesis holds, the principle is not necessarily valid in a more general case, as pointed out by Chung and Romano (2013).

In order to fill the existing gaps, we summarize our contributions in this chapter as follows. First, in Section 2.2.2, we prove Theorem 2.1 which shows that the difference between the critical values of the permutation KS test and the KS test defined in terms of the ecdfs of the two samples is asymptotically negligible for arbitrary underlying distributions of the two samples, which therefore applies to the difference between their p-values. This fundamental result provides the theoretical background in support of the recurrence method for computing p-values provided by formula (2.2.11) and its further enhancement given in formula (2.2.12).

We provide a formal description of the major steps of the method stemming from formula (2.2.11), with proofs and enlightening examples, given in Section 2.2.2 and Appendix 2.A, that are missing in the chapters of the above-mentioned authors. We have generalized the latter algorithm to allow for an arbitrary weight function (see Proposition 2.2 for the corresponding p-value). Such flexibility is important since, as known, the choice of weight function significantly affects the power of the test (see Examples 2.18 and 2.19, Appendix 2.C and also Büning, 2001). We have further enhanced the method, following Viehmann (2021), which leads to a slower but more stable and accurate computation of the KS p-values, as also shown in Section 4 and Appendix D. To match the tradeoff between speed and accuracy, we have implemented both versions of the algorithm in the R function `KS2sample`.

Second, in Section 2.2.3 we derive a closed form formula (c.f. (2.2.15), Theorem 2.5) for the asymptotic distribution, of the two-sample Kolmogorov-Smirnov test, which is valid for arbitrary samples, allowing ties in the observations. To the best of our knowledge, the asymptotic distribution of the two-sample KS test has not been investigated in the literature in the case of tied observations. Therefore, formula (2.2.15) and Theorem 2.5 represent a novel contribution of theoretical and numerical importance. We demonstrate (see Example 2.13 and Table 2.2) that the asymptotic formula (2.2.15) is a numerically efficient alternative to the exact method, for computing KS p-values when the sum of sample sizes exceeds $90,000$.

Third, in Section 2.2.5, we show how the method based on the recurrence formula (2.2.11) can be further generalized in order to compute exact Kuiper p-values allowing arbitrary continuous, discrete or mixed observations in the sample. To the best of our knowledge, the computation of exact Kuiper p-values has only been considered for the continuous case and equal sample sizes (Maag and Stephens, 1968; Hirakawa, 1973). We have also enhanced the computation in the spirit of Viehmann (2021), to improve its accuracy. To match the tradeoff between speed and accuracy, both of these versions of the algorithm for the KS and Kuiper p-values are implemented in the two functions `KS2sample` and `Kuiper2sample` and are fully investigated and compared in Section 2.4.

Fourth, we present a power comparison between the Kuiper test and the KS test with different weight functions. We show that the Kuiper test represents a more powerful alternative in the case of discrete observations, that seems not to have been previously investigated (see Appendix 2.C).

Lastly, we present a thorough review of the software that computes KS and Kuiper p-values, which is motivated by several reasons. Firstly, the majority of implementations of the KS and the Kuiper tests rely on approximate/resampling (rather than exact) methods to compute p-values. Secondly, it appears that the majority of the existing statistical software packages, (see Section 2.4 for a short summary and further details) have limited capabilities and give only p-values for the unweighted KS test, assuming single observations in the sample, not offering sufficient flexibility in choosing the weight function, which ultimately affects the power of the KS test (see Examples 2.18 and 2.19 in Appendix 2.C). Thirdly, al-

though the Kuiper test is applicable for circular observations with ties, the existing software for computing its p-values only work for the case of no ties, as pointed out by Boulesteix and Strobl (2007) and also evidenced by our thorough comparison in Section 2.4.2. However, goodness-of-fit testing on circular data with ties is also an important task, as confirmed by Jammalamadaka et al. (2020), that needs to be considered.

As mentioned, here we address all these deficiencies by developing the efficient and numerically stable R functions `KS2sample` and `Kuiper2sample`. We include the latter in the R package **KSgeneral**, previously developed by Dimitrova, Kaishev and Tan (2020). The latter package allows for computing p-values of the one-sample KS test for arbitrary continuous, discrete, or mixed null distribution. It has gained significant popularity among researchers and practitioners across various disciplines. This is evidenced by the large number of downloads of **KSgeneral** which exceeds 34,000 of present. The functions `KS2sample` and `Kuiper2sample` extend the capabilities of the **KSgeneral** to the two-sample goodness-of-fit testing with the KS and Kuiper tests. We have also developed a Mathematica code that computes KS and Kuiper p-values with arbitrary accuracy (albeit quite slow), and have used the latter as true benchmark p-values in the software comparisons that we present in Section 2.4 and Appendix 2.D. The results of the latter comparison are yet another contribution of this chapter that we believe is worth highlighting.

The chapter is organized as follows. In Section 2.2, we briefly introduce the weighted KS and Kuiper tests and describe the proposed methods for computing their corresponding p-values. In Theorem 2.1, we give a theoretical justification of the latter methods. In Section 2.2.3, we present formula (2.2.15) for the asymptotic distribution of the two-sample KS test. In Section 2.3, we give two examples, the first one illustrating the use of the R function `Kuiper2sample` on circular data with ties, for computing Kuiper p-values, and the second one illustrating the numerical performance of the asymptotic formula (2.2.15). Then in Section 2.4, we give a thorough overview of the existing software implementations for computing two-sample KS and Kuiper p-values and select some of them as benchmark implementations with whom we further compare our `KS2sample` and `Kuiper2sample` functions in terms of speed and accuracy. Finally, in Section 2.5, we summarize our findings. In

Appendix 2.A, we give some auxiliary results and proofs of the propositions from Section 2.2. In Appendix 2.B, we give some examples that illustrate the algorithms behind the functions `KS2sample` and `Kuiper2sample` and their numerical implementation. In Appendix 2.C, we present power comparisons between the KS and the Kuiper tests. Finally, Appendix 2.D contains numerical results that support the comparisons and conclusions of Section 2.4.

2.2 The Two-sample KS and Kuiper Test

2.2.1 The Two-sample KS Test

The two-sample KS statistic is designed to test whether two samples come from an unspecified common probability distribution. More precisely, let $F(x)$ and $G(x)$ be two cumulative distribution functions, either continuous or discrete (or mixed). Let $\boldsymbol{X}_m = (X_1, ..., X_m)$ and $\boldsymbol{Y}_n = (Y_1, ..., Y_n)$ be two mutually independent random samples drawn correspondingly from these two (unknown) distributions. We want to test the null hypothesis $H_0 : F(x) = G(x)$ for all x, either against the alternative hypothesis $H_1 : F(x) \neq G(x)$ for at least one x, which corresponds to the two-sided test, or against $H_1 : F(x) > G(x)$ and $H_1 : F(x) < G(x)$ for at least one x, which corresponds to the two one-sided tests. The two sample KS statistics that are used to test these hypotheses are generally defined as:

$$\begin{aligned} \Delta_{m,n} &= \sup_{x\in\mathbb{R}} |F_m(x) - G_n(x)| W(E_{m+n}(x)) \\ \Delta^+_{m,n} &= \sup_{x\in\mathbb{R}} \{F_m(x) - G_n(x)\} W(E_{m+n}(x)) \\ \Delta^-_{m,n} &= \sup_{x\in\mathbb{R}} \{G_n(x) - F_m(x)\} W(E_{m+n}(x)), \end{aligned} \tag{2.2.1}$$

where $\Delta_{m,n}$ denotes the two-sided test, $\Delta^+_{m,n}$ and $\Delta^-_{m,n}$ denote the two one-sided tests, $F_m(x)$ and $G_n(x)$ are the empirical (cumulative) distribution functions (edf) of $\boldsymbol{X}_m$ and $\boldsymbol{Y}_n$, $E_{m+n}(x)$ is the edf of the pooled sample $\boldsymbol{Z}_{m+n} := (X_1, ..., X_m, Y_1, ...Y_n)$, and $W(t)$ is a non-negative weight function, defined on $t \in [0, 1]$.

The choice of weight function in (2.2.1) leads to different specifications of the KS test statistic which, as noted in the introduction, influences its power. Following are some examples of choices of $W(t)$ for the two-sided KS test, $\Delta_{m,n}$ in (2.2.1). For instance, if $W(t) \equiv 1$, the latter test, denoted as $\Delta^0_{m,n}$ and referred to as the

unweighted KS test is given as:

$$\Delta^0_{m,n} = \sup_x |F_m(x) - G_n(x)|. \tag{2.2.2}$$

When $W(t) = 1/[t(1-t)]^{1/2}$, the KS test, denoted as $\Delta^{1/2}_{m,n}$, and referred to as the variance-stabilized weighted KS statistic (see Canner, 1975), is defined as:

$$\Delta^{1/2}_{m,n} = \sup_x \frac{|F_m(x) - G_n(x)|}{\sqrt{E_{m+n}(x)(1 - E_{m+n}(x))}}, \tag{2.2.3}$$

It is not difficult to see that the weight functions from (2.2.2) and (2.2.3) generalize to (see Finner and Gontscharuk, 2018):

$$W(t) = 1/[t(1-t)]^{\nu}, \tag{2.2.4}$$

where $0 \leq \nu \leq 1$. The latter defines a family of weighted KS tests, $\Delta^{\nu}_{m,n}$ $(0 \leq \nu \leq 1)$, which covers $\Delta^0_{m,n}$ and $\Delta^{1/2}_{m,n}$, for the choices, $\nu = 0$ and $\nu = 1/2$.

It has been pointed out by Finner and Gontscharuk (2018) that when the samples are drawn from two normal distributions, correspondingly with different mean and variance, the power of the statistic, $\Delta^{\nu}_{m,n}$, $\nu \in [0,1]$, as a function of ν, is approximately unimodal. The maximum power is achieved when $\nu \in [0.3, 0.6]$. When both the distributions of null and alternative hypothesis are left skewed and heavy-tailed, Büning (2001) suggested using the weight function $W(t) = 1/[t(2-t)]^{1/2}$ which ensures higher power than the test $\Delta^{1/2}_{m,n}$ (see also examples in Appendix 2.C).

2.2.2 Computing Exact P values of the KS Test

Let us note that since the distribution of $\Delta_{m,n}$ is unknown when F and G are two arbitrary distributions, computing its p-values is not directly possible. In order to overcome this difficulty, we consider an alternative formulation of the KS test, referred to as the permutation KS test. The latter is defined over the space of all possible pairs $C = \binom{m+n}{m}$ of samples $\tilde{\boldsymbol{X}}_m$ and $\tilde{\boldsymbol{Y}}_n$ of sizes m and n randomly drawn without replacement from the pooled sample $\boldsymbol{Z}_{m+n}$. We show in Theorem 2.1 that the difference between the critical value of the permutation KS test and that of $\Delta_{m,n}$ is asymptotically negligible for arbitrary F and G, which therefore applies to the difference between their p-values. This fundamental result provides the theoretical

background in support of the recurrence method for computing p-values provided by formula (2.2.11) and its further enhancement given in formula (2.2.12). Note that a recurrence method similar to that given in formula (2.2.11) but for the unweighted KS test has been independently considered by Nikiforov (1994), Hilton et al. (1994), and Schröer and Trenkler (1995), but without theoretical justification.

We also show that the formula (2.2.11) can be further simplified to the recurrence formula (2.2.12), which leads to an algorithm recently considered by Viehmann (2021). The latter is slower but more numerically stable and accurate, as demonstrated in Section 2.4. We have implemented both formulas (2.2.11) and (2.2.12) in the C++ functions `ks2sample_cpp`, `ks2sample_c_cpp`, the R function `KS2sample` and in Mathematica (Wolfram Research, Inc, 2024), which are more general than Nikiforov's Fortran 77 code, by allowing for arbitrary weight functions. It will be instructive to briefly highlight the main steps of the algorithm we have implemented.

For a particular realization of the pooled sample $\boldsymbol{Z}_{m+n}$, let there be k distinct values, $a_1 < a_2 < ... < a_k$, in the ordered, pooled sample $(z_1 \leq z_2 \leq \ldots \leq z_{m+n})$, where $k \leq m+n$. Denote by m_i the number of times a_i appears in the pooled sample, $i = 1, \ldots, k$. Given the pooled sample $\boldsymbol{Z}_{m+n}$, the p-value is then defined as the conditional probability

$$P = \mathbb{P}\left(D_{m,n} \geq q | \boldsymbol{Z}_{m+n}\right), \tag{2.2.5}$$

where $D_{m,n}$ is the permutation KS statistic, defined as in (2.2.1), for two samples $\tilde{\boldsymbol{X}}_m$ and $\tilde{\boldsymbol{Y}}_n$ of sizes m and n, randomly drawn from $\boldsymbol{Z}_{m+n}$ without replacement and $q \in [0,1]$. To simplify notation, we will further omit the condition in (2.2.5) and write $\mathbb{P}\left(D_{m,n} \geq q\right)$.

Formally, given the pooled sample $\boldsymbol{Z}_{m+n}$, the permutation KS test $D_{m,n}$ in (2.2.5) is defined as

$$D_{m,n}(\omega) = \sup_{x \in \mathbb{R}} |F_m(x,\omega) - G_n(x,\omega)| W(E_{m+n}(x)), \tag{2.2.6}$$

where $D_{m,n}$ is regarded as a mapping on the space Ω of all $C = \binom{m+n}{m}$ possible pairs of empirical distribution functions $(F_m(x,\omega), G_n(x,\omega))$, $\omega \in \Omega$. Each such pair corresponds to a pair of $\tilde{\boldsymbol{X}}_m$ and $\tilde{\boldsymbol{Y}}n$ obtained by randomly drawing (without re-

placement) from the pooled sample $\boldsymbol{Z}m+n$, as follows. First, m observations are drawn at random without replacement, forming the first sample $\tilde{\boldsymbol{X}}_m$, with corresponding edf, $F_m(x,\omega)$. The remaining n observations are then assigned to the second sample $\tilde{\boldsymbol{Y}}_n$, with corresponding edf $G_n(x,\omega)$. Resampling in this way will result in the occurrence of all the C possible pairs of edfs $F_m(x,\omega)$ and $G_n(x,\omega)$, $\omega \in \Omega$. The pairs of edf's may be coincident if there are ties in the data and each pair, $F_m(x,\omega)$ and $G_n(x,\omega)$ occurs with probability $1/C$ (see Example 2.3). The above-mentioned resampling procedure is a general permutation and randomization tests construction (See Lehmann and Romano, 2005, Section 15.2). Alternatively, following Section 3.8.1, van der Vaart and Wellner (2023), $F_m(x,\omega)$ and $G_n(x,\omega)$ can be equivalently viewed as the ecdfs of the permutation samples $\boldsymbol{X}_m^\dagger = \{Z_{R_1},\ldots,Z_{R_m}\}$ and $\boldsymbol{Y}_n^\dagger = \{Z_{R_{m+1}},\ldots,Z_{R_{m+n}}\}$, where $\boldsymbol{R}_{m+n} = (R_1,R_2,\ldots,R_{m+n})$ is independent from the samples $\boldsymbol{X}_m$ and $\boldsymbol{Y}_n$ and uniformly distributed on the set of permutations of $\{1,2,\ldots,m+n\}$, i.e. every possible realization of $\boldsymbol{R}_{m+n}$ has probability $1/(m+n)!$ (see (4.3.9) in Section 4.3.2).

Considering the conditional p-value P in (2.2.5) is important, since it allows us to show that the difference between the critical levels of $D_{m,n}$ and $\Delta_{m,n}$ becomes asymptotically negligible, which leads to the conditional p-value P converging to the unconditional p-value $\mathbb{P}(\Delta_{m,n} \geq q)$ as shown in the following theorem.

Theorem 2.1. *Assume that $\boldsymbol{X}_m$ and $\boldsymbol{Y}_n$ come correspondingly from the distributions $F(x)$ and $G(x)$. Denote by $q^\dagger_{m,n}(p)$ the critical level of the statistic $\sqrt{\frac{mn}{m+n}}D_{m,n}$, given $\boldsymbol{Z}_{m+n}$, i.e.*

$$q^\dagger_{m,n}(p) = \inf\{q : \mathbb{P}\left(\sqrt{\frac{mn}{m+n}}D_{m,n} \geq q\right) \leq p|\boldsymbol{Z}_{m+n}\}.$$

Denote by $q^*_{m,n}(p)$ the critical value of $\sqrt{\frac{mn}{m+n}}\Delta_{m,n}$ when $\boldsymbol{X}_m$ and $\boldsymbol{Y}_n$ come from the distribution $\eta F + (1-\eta)G$, $\eta \in (0,1)$, i.e.

$$q^*_{m,n}(p) = \inf\{q : \mathbb{P}\left(\sqrt{\frac{mn}{m+n}}\Delta_{m,n} \geq q\right) \leq p\}.$$

When the weighted function W in $\Delta_{m,n}$ is bounded and uniformly continuous and for a fixed p, we have

$$q^\dagger_{m,n}(p) - q^*_{m,n}(p) \xrightarrow{\mathbb{P}} 0,$$

as $m,n \to \infty$ *and* $\frac{m}{m+n} \to \eta$, *where* $\xrightarrow{\mathbb{P}}$ *stands for convergence in probability.*

The normalizing factor $\sqrt{\frac{mn}{m+n}}$ ensures that the convergence result is non-trivial, since $q^{\dagger}_{m,n}(p)$ does not converge to 0. When both F and G are continuous, the result is less interesting, since both $\Delta_{m,n}$ and $D_{m,n}$ are distribution-free. However, it is particularly important when both F and G are discontinuous cdfs and the null hypothesis $F = G$ may fail to hold.

Let us note that a result similar to Theorem 2.1 may not hold for an arbitrary two-sample statistic and its corresponding permutation test. This is illustrated by Example 5.3 of Chung and Romano (2013). Let us also note that a similar asymptotic consistency result for the unweighted KS test is provided by van der Vaart and Wellner (2023) using a different approach. Theorem 2.1 is more general as it covers the weighted KS test and naturally leads to the numerical algorithm for computing KS p-values that follows.

In order to compute the p-value P in (2.2.5), we alternatively express it as,

$$P = 1 - \frac{N}{C}, \tag{2.2.7}$$

where, N is the number of pairs, $F_m(x,\omega)$ and $G_n(x,\omega)$, drawn from the pooled sample $\boldsymbol{Z}_{m+n}$, for which $D_{m,n} < q$. So in order to compute the p-value, P, it suffices to find N.

In order to find N, let us introduce the integer-valued grid $R = \{(i,j) : 0 \leq i \leq m, \quad 0 \leq j \leq n\}$. Define a trajectory on R as a point sequence $\{i_l, j_l\}_{l=0}^{m+n}$ that satisfies the following conditions

$$\begin{aligned} &1. \quad (i_{l+1}, j_{l+1}) = (i_l + 1, j_l) \quad \text{or} \quad (i_{l+1}, j_{l+1}) = (i_l, j_l + 1) \quad (0 \leq l \leq m+n-1) \\ &2. \quad (i_0, j_0) = (0,0) \quad \text{and} \quad (i_{m+n}, j_{m+n}) = (m,n) \end{aligned} \tag{2.2.8}$$

A trajectory moves to the right from (i_l, j_l) to $(i_l + 1, j_l)$ if z_l, the l-th value in $\boldsymbol{Z}_{m+n}$, is from the first sample and it moves up to $(i_l, j_l + 1)$ otherwise. Let us note that, a unique trajectory in R corresponds to each of the C pairs of edfs, $F_m(x,\omega)$ and $G_n(x,\omega)$. As we show in Appendix 2.A, the number N is calculated using the following proposition.

Proposition 2.2. *The number N, of pairs $F_m(x,\omega)$ and $G_n(x,\omega)$ for which $D_{m,n} < q$, coincides with the number of trajectories that lie wholly in a subset, S of R, such that,*

$$S = R_1 \cup R_2 \cup ... \cup R_k, \tag{2.2.9}$$

where R_l, $l = 1,...,k$, is a set of points $(i,j) \in R$, such that

$$\begin{aligned} &\min\{i : d(i,j) < q, i+j = T_{l-1}\} \leq i \leq \max\{i : d(i,j) < q, i+j = T_l\}, \\ &\min\{j : d(i,j) < q, i+j = T_{l-1}\} \leq j \leq \max\{j : d(i,j) < q, i+j = T_l\} \end{aligned} \tag{2.2.10}$$

where $T_l = m_1 + ... + m_l$, $T_0 = 0$ and $d(i,j)$ is defined as shown in Table 2.1, depending on the particular definition of the KS test, $W(t)$ is the weight function defined in (2.2.1).

Table 2.1: Value of $d(i,j)$ depending on the KS test.

	Unweighted	Weighted
One-sided	$\pm(i/m - j/n)$	$\pm(i/m - j/n)W((i+j)/(m+n))$
Two-sided	$\lvert i/m - j/n\rvert$	$\lvert i/m - j/n\rvert W((i+j)/(m+n))$

Proposition 2.2 is a generalization of formula (5) in Nikiforov (1994). The latter applies to the unweighted test, in which case S is determined only by the value of q. In more general cases, i.e., for the weighted KS test with arbitrary weight function, the subset S in (2.2.9) is determined not only by the value of q but also by the form of $d(i,j)$ which is specified by the corresponding KS test. Once S is determined, N is obtainable by counting the total number of trajectories that lie in the subset S. For a better understanding of the rationale behind the subset S, see Examples 2.3 and 2.4 and the proof of Proposition 2.2 provided in Appendix 2.A.

Following Proposition 2.2, the subset, S, defined in (2.2.9) is then used to calculate the number, N, (c.f (2.2.7)), as follows. Denote by $B_S(i,j)$, the total number of ways to move from point $(0,0)$ to point (i,j) $(1 \leq i \leq m, 1 \leq j \leq n)$ while staying strictly inside the region S. Since the number N coincides with $B_S(m,n)$, i.e., $B_S(m,n) = N$ by definition, it suffices to find $B_S(m,n)$, in order to find the p-value $P = 1 - N/C$. For the purpose, $B_S(i,j)$ is computed for each point $(i,j) \in R$,

by the following recurrence formula

$$B_S(i,j) = \mathbb{1}_S(i,j)[B_S(i,j-1) + B_S(i-1,j)] \text{ for } (i \geq 1, j \geq 1), \tag{2.2.11}$$

with starting values $B_S(0,j) = \mathbb{1}_S(0,j)$ and $B_S(i,0) = \mathbb{1}_S(i,0)$, where $\mathbb{1}_S(\cdot)$ denotes the indicator function. Following (2.2.11), an algorithm containing a loop with index l is designed to simultaneously calculate all the values of $B_S(i,j)$ with $i+j = l+1$, using all the $B_S(i,j)$ with $i+j = l$. When implementing in C++ and R, appropriate scaling is applied to ensure that $B_S(i,j)$ does not become too large, which may lead to possible loss of accuracy and numerical stability. We provide Examples 2.3 and 2.4 illustrating this algorithm as follows.

Example 2.3. For simplicity, we illustrate the algorithm on the example of the unweighted KS test given in (2.2.2), for the case, $W(t) = 1$. Given two samples, $\boldsymbol{X} = \{10,30,30\}$ and $\boldsymbol{Y} = \{30,40,40,50\}$, of sizes, $m = 3$ and $n = 4$, for which the (observed) unweighted, two-sided KS statistic $q = 0.75$. The four ordered, distinct values are $(a_1,a_2,a_3,a_4) = (10,30,40,50)$ with number of repetitions $m_1 = 1, m_2 = 3, m_3 = 2, m_4 = 1$, in the pooled sample $\boldsymbol{Z} = (10,30,30,30,40,40,50)$, whose entries can be labeled as $10(z_1)$, $30(z_2)$, $30(z_3)$, $30(z_4)$, $40(z_5)$, $40(z_6)$, and $50(z_7)$.

Following the re-sampling scheme described above, the number, C of all pairs of samples and their corresponding trajectories in R, that can be drawn from $\boldsymbol{Z}$, is $C = \binom{7}{3} = 35$. It is not difficult to see that, only 11 out of the 35 pairs are distinct. The pairs and the numbers, r_i, $i = 1,\ldots,11$ of their repetitions are:

$$\{\{10,30,30\},\{30,40,40,50\}\}, r_1 = 3, \{\{10,30,40\},\{30,30,40,50\}\}, r_2 = 6,$$

$$\{\{10,30,50\},\{30,30,40,40\}\}, r_3 = 3, \{\{10,40,40\},\{30,30,30,50\}\}, r_4 = 1,$$

$$\{\{10,40,50\},\{30,30,30,40\}\}, r_5 = 2, \{\{30,30,30\},\{10,40,40,50\}\}, r_6 = 1,$$

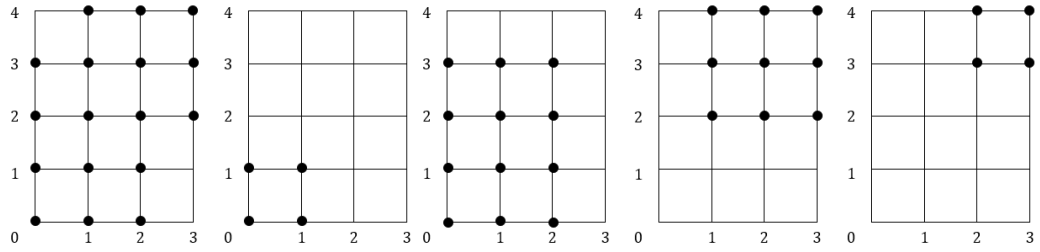
$$\{\{30,30,40\},\{10,30,40,50\}\}, r_7 = 6, \{\{30,30,50\},\{10,30,40,40\}\}, r_8 = 3,$$

$$\{\{30,40,40\},\{10,30,30,50\}\}, r_9 = 3, \{\{30,40,50\},\{10,30,30,40\}\}, r_{10} = 6,$$

$$\{\{40,40,50\},\{10,30,30,30\}\}, r_{11} = 1,$$

where $\sum_{i=1}^{11} r_i = 35$.

The set, S of all trajectories, for which $D_{m,n} < q$, obtained applying (2.2.9) and (2.2.10), is illustrated in Fig 2.1 (a), together with the subsets $R_i, i = 1,2,3,4$.



(a) Region S **(b)** Subset R_1 **(c)** Subset R_2 **(d)** Subset R_3 **(e)** Subset R_4

Figure 2.1: Region S and its Component when $D^0_{m,n} < 0.75$ with $(M_1, M_2, M_3, M_4) = (1,3,2,1)$.

As noted, a unique trajectory in R, corresponds to each of the 35 pairs of samples. To illustrate this, let us take the first distinct pair, $\{\{10,30,30\},\{30,40,40,50\}\}$, repeated $r_1 = 3$ times, for which $D_{m,n} = 0.75 = q$. Using the labels $Z_i, i = 1,\ldots,7$ of the observations in the pooled sample, these three pairs are

$$\{\{10(z_1),30(z_2),30(z_3)\},\{30(z_4),40(z_5),40(z_6),50(z_7)\}\},$$

$$\{\{10(z_1),30(z_2),30(z_4)\},\{30(z_3),40(z_5),40(z_6),50(z_7)\}\},$$

$$\{\{10(z_1),30(z_4),30(z_3)\},\{30(z_2),40(z_5),40(z_6),50(z_7)\}\}.$$

with trajectories,

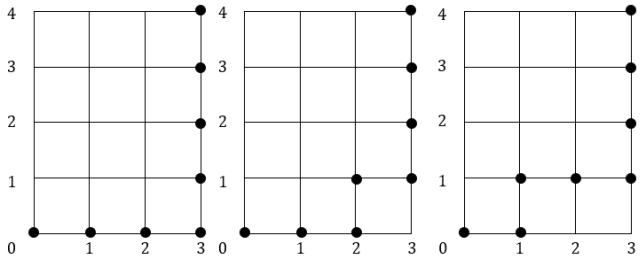
$$\{(0,0),(1,0),(2.0),(3,0),(3,1),(3,2),(3,3),(3,4)\},$$

$$\{(0,0),(1,0),(2,0),(2,1),(3,1),(3,2),(3,3),(3,4)\},$$

$$\{(0,0),(1,0),(1,1),(2,1),(3,1),(3,2),(3,3),(3,4)\},$$

illustrated in Figure 2.2 (a),(b) and (c), respectively.

From (2.2.10), we have that, $T_1 = 1, T_2 = 4, T_3 = 6, T_4 = 7$ and one can directly check that (2.A.11) gives the value $D_{m,n} = 0.75$, which coincides with the (observed) value $q = 0.75$, obtained from the edfs of the distinct pair



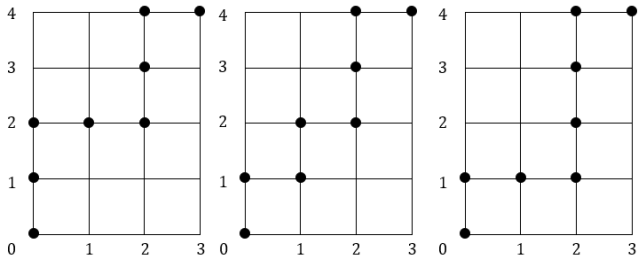
(a) Trajectory 1 **(b)** Trajectory 2 **(c)** Trajectory 3

Figure 2.2: Possible trajectories in the grid R with ties (Example 2.3).

$\{\{10,30,30\},\{30,40,40,50\}\}$.

In this particular example, under the condition that $D^0_{m,n} < 0.75$, (2.2.9) and (2.2.10) settle the subset S in lattice R, the shape of which is shown in Figure 2.1a. By further applying (2.2.11), it is not hard to find the fact that N, the number of pairs of samples for which $D^0_{m,n} < 0.75$, is equal to 30, which could also be found by exhausting all the possible ways of draws.

We provide another example to illustrate Proposition 2.2.



(a) Trajectory 1 **(b)** Trajectory 2 **(c)** Trajectory 3

Figure 2.3: Possible trajectories in grid R with ties (Example 2.4).

Example 2.4. Given two samples, $\boldsymbol{X} = \{30,30,50\}$ and $\boldsymbol{Y} = \{10,30,40,40\}$, drawing from the pooled sample $\boldsymbol{Z}$ in Example 2.3, for which the unweighted two-sided KS statistic $q = 1/3 < 0.75$. Applying the same approach as in Example 2.3, we find that there are three possible corresponding trajectories, as shown in Figure 2.3. Furthermore, the shape of S in Figure 2.1a shows that all 3 corresponding trajectories are subsets of S.

Computing the p-value based on the recursion formula (2.2.11) can be further optimized to improve accuracy and numerical stability, in the spirit of Viehmann (2021). To achieve this, we alternatively define the number $J_S(i,j)$ as the proportion

of the trajectories from (0,0) to (i,j) that do not fully stay inside of the subset S. Then the p-value P is directly expressed as $P = J_S(m,n)$. In addition, it is not difficult to see that $J_S(i,j) = 1 - B_S(i,j)/\binom{i+j}{i}$. Hence, one can further rewrite (2.2.11) as the recursion formula:

$$J_S(i,j) = 1 + \mathbb{1}_S(i,j)[\frac{i}{i+j}J_S(i-1,j) + \frac{i}{i+j}J_S(i,j-1) - 1] \text{ for } (i \geq 1, j \geq 1) \tag{2.2.12}$$

with starting values $J_S(0,j) = 1 - \mathbb{1}_S(0,j)$ and $J_S(i,0) = 1 - \mathbb{1}_S(i,0)$. To implement (2.2.12), an algorithm containing a loop with index l is designed to compute $J_S(i,j)$ with $i+j = l+1$. Note that each step of the algorithm would only involve computing the weighted averages of floating numbers between 0 and 1, hence there is no loss of accuracy in each step, which ensures the accuracy of P.

2.2.3 The Asymptotic Distribution of $D_{m,n}$

It is well known that when samples are assumed from continuous distributions with cdf F and G, $\Delta_{m,n}$ is distribution-free under the null hypothesis. When $m,n \to \infty$ and $m/(m+n) \to \lambda \in (0,1)$, the asymptotic distribution of the unweighted $\Delta^0_{m,n}$ (also $D^0_{m,n}$) coincides with the famous Kolmogorov distribution,

$$\mathbb{P}\left\{\sqrt{\frac{nm}{n+m}}\Delta^0_{m,n} \leq x\right\} = 1 - 2\sum_{k=1}^{\infty}(-1)^{k-1}e^{-2k^2x^2} \tag{2.2.13}$$

When the underlying null distributions $F = G$ are purely discrete, Eplett (1982) shows that, the unweighted $\Delta^0_{m,n}$ weekly converge to a Gaussian process. While this result is theoretically useful, it is less appealing for numerical purposes, since in reality, we do not have any knowledge of F or G and whether $F = G$. Therefore, the distribution of $\Delta^0_{m,n}$ is unobservable. The general case when F and G may be continuous, discrete or mixed and $F \neq G$ seems not to have been considered jointly in the literature, and our goal here will be to fill in this gap and provide a closed-form expression for the asymptotic distribution of the unweighted permutation test $D^0_{m,n}$. The proof of our formula given in Theorem 2.5 is based on several results. First, we recall the results of van der Vaart and Wellner (2023), who show that the two-sample and one-sample KS tests converge to a Brownian bridge. Second, based on the latter convergence, we show that the explicit formula of Dimitrova et al. (2020) for the

asymptotic distribution of the unweighted one-sample KS test, assuming arbitrary null distribution, applies also to the two-sample case (see Theorem 2.5 and its proof in Appendix 2.A).

Let us assume that $m/(m+n) \to \eta \in (0,1)$ when $m,n \to \infty$. Denote by $E(x) = \eta F(x) + (1-\eta)G(x)$. By Theorem 2.1, when $\boldsymbol{X}_m$ and $\boldsymbol{Y}_n$ come from F and G, the pooled sample $\boldsymbol{Z}_{m+n} = (X_1,\ldots,X_m,Y_1,\ldots,Y_n)$ in the unweighted KS test $D_{m,n}$ behaves as if it directly comes from the pooled distribution E.[1] Clearly, any assumption on the jump structure of F and G is valid for the jump structure of E. Therefore, we can assume that, E has a jump structure and shape, as defined in Dimitrova et al. (2020) for F in the one-sample case. More precisely, assume that E has finite number of jumps Λ, occurring at points x_l, $l = 1,\ldots,\Lambda$, with $E(x_l-) = f_{2l-1}$ and $E(x_l) = f_{2l}$. And we distinguish between increasing segments, i.e. $f_{2l-2} < f_{2l-1}$, and flat segments, i.e. $f_{2l-2} = f_{2l-1}$. Additionally, we set $f_0 = 0$ and $f_{2\Lambda+1} \equiv 1$ to complete the sequence.

Denote by $\nu_1,\nu_2,\ldots$ the number of increasing segments appearing consecutively, and by $\omega_1,\omega_2,\ldots$ the number of flat segments appearing consecutively. Without loss of generality, we assume there are p groups of increasing segments and flat segments, i.e. $\Lambda = \nu_1+\omega_1+\cdots+\nu_p+\omega_p$, which allows all the jumps x_l to appear in the following order

$$\begin{aligned}\{&x_1,\ldots,x_{\nu_1},x_{\nu_1+1},\ldots x_{\nu_1+\omega_1},x_{\nu_1+\omega_1+1},\ldots,x_{\nu_1+\omega_1+\nu_2},x_{\nu_1+\omega_1+\nu_2+1},\ldots,\\ &x_{\nu_1+\omega_1+\nu_2+\omega_2},\ldots,x_{\nu_1+\omega_1+\cdots+\omega_{p-1}+1},\ldots,x_{\nu_1+\omega_1+\cdots+\omega_{p-1}+\nu_p},\\ &x_{\nu_1+\omega_1+\cdots+\omega_{p-1}+\nu_p+1},\ldots x_{\nu_1+\omega_1+\cdots+\omega_{p-1}+\nu_p+\omega_p}\}\end{aligned} \tag{2.2.14}$$

where $\omega_1 \geq 0$, $\nu_1 \geq 0$, $\omega_1+\nu_1 > 0$; $\nu_l > 0$, $l = 2,\ldots,p$, $\omega_l > 0$, $l = 2,\ldots,p-1$; $\omega_p \geq 0$.

Based the above framework, we now give the asymptotic distribution as follows:

Theorem 2.5. *Given the realization of $\boldsymbol{X}_m$ and $\boldsymbol{Y}_n$ coming from F and G respectively, denote by $\Phi(x)$ the limiting distribution of $\mathbb{P}(\sqrt{\frac{nm}{n+m}}D^0_{m,n} \leq x)$ when $m,n \to \infty$ with $m/(m+n) \to \eta \in (0,1)$ and $W(\cdot) \equiv 1$. If the joint distribution $E(x) = \eta F(x) + (1-\eta)G(x)$ has a structure of jumps as in (2.2.14), under the null*

[1] This consideration relies on Theorem 2.1, and it fails to hold in a more general case. See more discussion in Chung and Romano (2013).

hypothesis, when $f_{2\Lambda} = f_{2\Lambda+1}$,

$$\Phi(x) = \sum_{j_1=-\infty}^{\infty} \cdots \sum_{j_p=-\infty}^{\infty} \left((-1)^{j_1+\cdots+j_p}\right) c \int_{-x}^{x} \cdots \int_{-x}^{x} e^{\psi} dz_1 \cdots dz_{2v_p+w_p-1}, \quad (2.2.15)$$

where

$$\begin{aligned} c = \prod_{i=1}^{p} \Bigg(\prod_{l=1}^{\nu_i} & (f_{2(v_{i-1}+w_{i-1}+l)-1} - f_{2(v_{i-1}+w_{i-1}+l)-2})^{-1/2} \\ & (f_{2(v_{i-1}+w_{i-1}+l)} - f_{2(v_{i-1}+w_{i-1}+l)-1})^{-1/2} \Bigg) \\ & \times \left(\prod_{l=1}^{\omega_i} \left(f_{2(v_i+w_{i-1}+l)} - f_{2(v_i+w_{i-1}+l)-1} \right)^{-1/2} \right) (2\pi)^{-\frac{2v_p+w_p-1}{2}}, \end{aligned} \quad (2.2.16)$$

and

$$\begin{aligned} \psi = -\frac{1}{2} \sum_{i=1}^{p} \Bigg\{ \sum_{l=1}^{\nu_i} \Bigg[& \frac{(z_{2(v_{i-1}+l)+w_{i-1}} - z_{2(v_{i-1}+l)+w_{i-1}-1})^2}{f_{2(v_{i-1}+w_{i-1}+l)} - f_{2(v_{i-1}+w_{i-1}+l)-1}} \\ & + \frac{(z_{2(v_{i-1}+l)+w_{i-1}-1} - (-1)^{j_{(v_{i-1}+l)}} z_{2(v_{i-1}+l)+w_{i-1}-2} - 2x j_{(v_{i-1}+l)})^2}{f_{2(v_{i-1}+w_{i-1}+l)-1} - f_{2(v_{i-1}+w_{i-1}+l)-2}} \Bigg] \\ & + \sum_{l=1}^{\omega_i} \left[\frac{(z_{2v_i+w_{i-1}+l} - z_{2v_i+w_{i-1}+l-1})^2}{f_{2(v_i+w_{i-1}+l)} - f_{2(v_i+w_{i-1}+l)-1}} \right] \Bigg\}, \end{aligned} \quad (2.2.17)$$

with $\nu_0 = \omega_0 = 0; \nu_0 = w_0 = 0; v_i = \sum_{k=1}^{i} \nu_k; w_i = \sum_{k=1}^{i} \omega_k, v_p + w_p = \Lambda$, *and* $z_0 = z_{2v_p+w_p} = 0$.

Furthermore, when $f_{2\Lambda} < f_{2\Lambda+1}$ *and that* $v_p + w_p = \Lambda$, *one only need to substitute* c' *for* c *and* ψ' *for* ψ *in (2.2.15) to compute* $\Phi(x)$, *where*

$$c' = c(f_{2\Lambda+1} - f_{2\Lambda})^{-1/2}(2\pi)^{-1/2} \text{ and } \psi' = \psi + \frac{(-(-1)^{j_{v_p+1}} z_{2v_p+w_p} - 2x j_{v_p+1})^2}{f_{2\Lambda+1} - f_{2\Lambda}} \quad (2.2.18)$$

Corollary 2.6. *When* $E(x)$ *is purely discrete with jumps* Λ, *the limiting distribution* $\Phi(x)$ *in (2.2.15) becomes:*

$$\Phi(x) = (2\pi)^{-\frac{\Lambda-1}{2}} \prod_{l=1}^{\Lambda} (f_{2l} - f_{2l-1})^{-\frac{1}{2}} \int_{-x}^{x} \cdots \int_{-x}^{x} \exp\left[-\frac{1}{2} \left(\sum_{l=1}^{\Lambda} \frac{(z_l - z_{l-1})^2}{f_{2l} - f_{2l-1}} \right) \right] dz_1 \cdots dz_{\Lambda-1}. \quad (2.2.19)$$

Note that in view of Theorem 2.1, (2.2.19) also applies to the limiting distribution of $\Delta^0_{m,n}$, it is a closed-form explicit alternative to the result of Eplett (1982),

who only states convergence to a Gaussian process.

Remark 2.7. Let us note that, as long as $\Lambda = o((m+n)^{-\frac{1}{2}})$, the jump structure f_l, $l = 1,\ldots,2\Lambda$, of E can be directly estimated from the pooled sample $\boldsymbol{Z}_{m+n}$ as the corresponding frequency, so Theorem 2.5 and Corollary 2.6 can be applied to compute p-values of $D_{m,n}$, for large m and n. This is illustrated in Example 2.13, Section 2.3. In addition, for a fixed x, since $\frac{\partial}{\partial f_i}\Phi(x)$ are bounded and the estimators $\hat{f}_l$ converge to f_l at the rate $O_p((m+n)^{-\frac{1}{2}})$ for $l = 1,\ldots,2\Lambda$, the error $|\Phi(x) - \hat{\Phi}(x)|$ introduced by estimation is also of order $O_p((m+n)^{-\frac{1}{2}})$, where $\hat{\Phi}(x)$ is computed by (2.2.15) based on $\hat{f}_l$, $l = 1,\ldots,2\Lambda$.

2.2.4 The Two-sample Kuiper Test

The (unweighted) two sample Kuiper statistic (Kuiper, 1960) can also be used to test the null hypothesis $H_0 : F(x) = G(x)$ for all x, against the alternative hypothesis $H_1 : F(x) \neq G(x)$ for at least one x. It is defined as

$$\varsigma_{m,n} = \sup_{x\in\mathbb{R}}[F_m(x) - G_n(x)] - \inf_{x\in\mathbb{R}}[F_m(x) - G_n(x)], \tag{2.2.20}$$

and can be alternatively expressed as

$$\varsigma_{m,n} \equiv \Delta^{0+}_{m,n} + \Delta^{0-}_{m,n}, \tag{2.2.21}$$

where $\Delta^{0+}_{m,n}$ and $\Delta^{0-}_{m,n}$ are the one-sided two-sample KS tests defined in (2.2.1). As shown by Kuiper (1960), the Kuiper statistic is invariant to cyclic transformations, in particular, invariant under all shifts and parametrizations on the circle (as shown in Example 2.12). It makes the Kuiper test ideal for examining pairs of circular and seasonal data. Furthermore, this property guarantees that the Kuiper test is equally sensitive across the entire support of the distribution, i.e., as sensitive in the tails as around the median (Press et al., 2007). The property of uniform sensitivity is also supported by various numerical studies, which have pointed out that, the Kuiper test has equal or higher power than the two-sample Kolmogorov-Smirnov, Cramer-Von mises and Anderson-Darling tests in some cases, for instance, when the first sample comes from a Normal distribution and the second sample comes from the scale shifted first distribution (Foutz and Birch, 1982), when the first sample from a Normal and the second from a mixture of normal (Dowd 2020 and Wyłomańska et al.

2020). Lemeshko et al. (2014) have also shown that in testing simple hypotheses, the Kuiper and Watson goodness-of-fit tests have an advantage in power over the Kolmogorov–Smirnov, Cramer–von Mises, and Anderson–Darling tests.

2.2.5 Computing Exact P values of the Kuiper Test

As highlighted in (2.2.21), the Kuiper test is closely related to the two-sample KS test. Hence, to calculate the p-value of the (unweighted) two-sample Kuiper test, we need to extend the recursion formulas (2.2.11) and (2.2.12) for computing the p-value of the KS test. Note that this step is also a generalization of the approaches by Maag and Stephens (1968) and Hirakawa (1973), which only apply to equally large data samples without ties. In this section, we will briefly introduce the main procedure behind the proposed method.

For a particular realization of an ordered pooled sample $\boldsymbol{Z}_{m+n}$, similarly as in (2.2.5), the p-value for the permutation Kuiper test is defined as the probability

$$P' = \mathbb{P}\left(V_{m,n} \geq q | \boldsymbol{Z}_{m+n}\right), \tag{2.2.22}$$

where $V_{m,n}$ is the permutation Kuiper statistic, defined as in (2.2.20), on the space Ω, for two samples $\tilde{\boldsymbol{X}}_m$ and $\tilde{\boldsymbol{Y}}_n$ of sizes m and n, randomly drawn from $\boldsymbol{Z}_{m+n}$ without replacement, and $q \in [0,2]$.

Once again, the p-value of the permutation Kuiper test is asymptotically consistent with that of the Kuiper test, since we can also establish the result of consistency between the critical values of the Kuiper and its permutation statistics, similar to Theorem 2.1.

Theorem 2.8. *Theorem 2.1 remains valid if one correspondingly substitute $\Delta_{m,n}$ and $D_{m,n}$ with $\varsigma_{m,n}$ and $V_{m,n}$.*

Following the same logic as in (2.2.7), the p-value in (2.2.22) can be expressed as $P' = 1 - \frac{N'}{C}$, where N' is the number of pairs, $F_m(x,\omega)$ and $G_n(x,\omega)$, drawn from the pooled sample $\boldsymbol{Z}_{m+n}$, for which $V_{m,n} < q$. And in order to compute the p-value, P', it suffices to find N', which is given in the following proposition.

Proposition 2.9. *The number N', of pairs $F_m(x,\omega)$ and $G_n(x,\omega)$ for which $V_{m,n} <$*

q is calculated as:

$$N' = \begin{cases} \sum_{i=1}^{\lceil qr \rceil} N(\frac{i}{r}, \frac{\lceil qr \rceil - i + 1}{r}) - \sum_{i=1}^{\lceil qr \rceil - 1} N(\frac{i}{r}, \frac{\lceil qr \rceil - i}{r}) & q > 1/r \\ N(\frac{1}{r}, \frac{1}{r}) & q \leq 1/r \end{cases} \tag{2.2.23}$$

where $r = lcm(m,n)$ is the least common multiple of m and n, $N(a,b)$ represents the total number of pairs $F_m(x,\omega)$ and $G_n(x,\omega)$, drawn from the pooled sample $\boldsymbol{Z}_{m+n}$, for which $D^{0+}_{m,n} < a$ and $D^{0-}_{m,n} < b$, $D^{0+}_{m,n}$ and $D^{0-}_{m,n}$ are the one-sided unweighted KS tests, defined on the space Ω, and a and b take values $\{0, \frac{1}{r}, \frac{2}{r}, \ldots, 1\}$. In particular, $N(a,b)$ coincides with the number of trajectories that lie wholly in a subset, $S(a,b)$ of R, defined as

$$S(a,b) = R_1(a,b) \cup R_2(a,b) \cup \ldots \cup R_k(a,b), \tag{2.2.24}$$

where $R_l(a,b)$, $l = 1,\ldots,k$, is a set of points $(i,j) \in R$, such that

$$\begin{aligned} &\min\{i : (i/m - j/n) < a, i + j = T_{l-1}\} \leq i \leq \max\{i : (j/n - i/m) < b, i + j = T_l\}, \\ &\min\{j : (j/n - i/m) < b, i + j = T_{l-1}\} \leq j \leq \max\{j : (i/m - j/n) < a, i + j = T_l\} \end{aligned} \tag{2.2.25}$$

Remark 2.10. When $1 < q < 2$, the alternative for calculating the number N' is

$$N' = \sum_{i=\lceil qr \rceil - r}^{r+1} N(\frac{i}{r}, \frac{\lceil qr \rceil - i + 1}{r}) - \sum_{i=\lceil qr \rceil - r}^{r} N(\frac{i}{r}, \frac{\lceil qr \rceil - i}{r}), \tag{2.2.26}$$

which is a simplification of (2.2.23), and therefore leads to a more efficient calculation of the p-value when $q > 1$.

Therefore, to obtain N', which is then used for computing the p-value $P' = 1 - N'/C$ for the Kuiper test, we could compute the numbers $N(a,b)$ in (2.2.23) via the number $B_S(m,n)$ in (2.2.11) substituting S with $S(a,b)$. The algorithm is designed to have two separate loops. In the first loop, for a fixed index i, the values $N(\frac{i}{r}, \frac{\lceil qr \rceil - i + 1}{r})$ and $N(\frac{i}{r}, \frac{\lceil qr \rceil - i}{r})$ are computed. In the second loop, N' is then computed applying the first line of (2.2.23). In the C++ and R implementation, appropriate scaling is applied to these two loops, which leads to a loss of accuracy as in the case of KS test.

As in the case of the KS test, we further optimize the accuracy of the algorithm, following the idea of Viehmann (2021). To do so, we define $P(a,b)$ as the probability of $D^{0+}_{m,n} < a$ and $D^{0-}_{m,n} < b$, i.e., $P(a,b) = 1 - N(a,b)/C$. Then, (2.2.23) could be rewritten as a formula of the p-value P'

$$P' = \begin{cases} \sum_{i=1}^{\lceil qr \rceil} P(\frac{i}{r}, \frac{\lceil qr \rceil - i + 1}{r}) - \sum_{i=1}^{\lceil qr \rceil - 1} P(\frac{i}{r}, \frac{\lceil qr \rceil - i}{r}) & q > 1/r \\ P(\frac{1}{r}, \frac{1}{r}) & q \leq 1/r \end{cases} \tag{2.2.27}$$

Note that each value $P(a,b)$ in (2.2.27) can be directly computed from (2.2.12) using the number $J_S(m,n)$, substituting the set S with $S(a,b)$, defined in (2.2.24). Therefore, the algorithm contains one loop with respect to one index i, for computing $P(\frac{i}{r}, \frac{\lceil qr \rceil - i + 1}{r})$ and $P(\frac{i}{r}, \frac{\lceil qr \rceil - i}{r})$.

Remark 2.11. Further generalizations to the weighted two-sample Kuiper tests are computational intensive, since $D^{0+}_{m,n}$ and $D^{0-}_{m,n}$ may not be the integer multiples of $1/r$, which is the condition required for deriving (2.2.23) from (2.A.15).

We have implemented our more general approach in the C++ function `kuiper2sample_cpp`, `kuiper2sample_c_cpp`, the R function `Kuiper2sample` and in Mathematica, that applies to arbitrary data samples, possibly with ties.

2.3 Numerical Implementation and Examples

As mentioned in Sections 2.2.2 and 2.2.5, we have implemented the proposed efficient and exact calculation of p-values in C++, R and Mathematica (Wolfram Research, Inc, 2024). The Mathematica implementation, following (2.2.11) and (2.2.23), allows for computing the KS and Kuiper p-values with arbitrary precision, but at a relatively high computational cost (see Appendix 2.D), especially when sample sizes are very large. We use the latter p-values as the true values for comparison purposes. We have implemented (2.2.12) and (2.2.11) correspondingly in the C++ functions `ks2sample_cpp` and `ks2sample_c_cpp`, and (2.2.27) and (2.2.23) correspondingly in the functions `kuiper2sample_cpp` and `kuiper2sample_c_cpp`. The R function `KS2sample` wraps the C++ functions `ks2sample_cpp` and `ks2sample_c_cpp` , whereas the R function `Kuiper2sample` wraps the C++ functions `kuiper2sample_cpp` and `kuiper2sample_c_cpp`. We illustrate the use of `KS2sample` and `Kuiper2sample` in Appendix 2.B.

As illustrated in Section 2.4 and Appendix 2.D, the R functions `KS2sample` and `Kuiper2sample` yield p-values with high precision and very small run times, compared with other implementations in the main-stream statistical software, which makes these functions preferable for practical use. For example, the function `ks2sample_cpp` computes KS p-values following (2.2.12) with at least 14 but typically 16–17 correct digits for sample sizes $m + n \leq 100,000$. The function `kuiper2sample_cpp` provides Kuiper p-values with at least 7 and up to 14 correct digits (see Section 2.4).

As we highlighted, the Kuiper test is also applicable for testing circular data, which is illustrated in the following example.

Example 2.12. Assume we have two discrete samples `sample3` and `sample4` observed on a circle, containing angular observations in degrees or radians. The `sample3` and `sample4` contain 120 and 149 observations respectively, with all the observations equal to one of four angles, as shown in Figure 2.4.

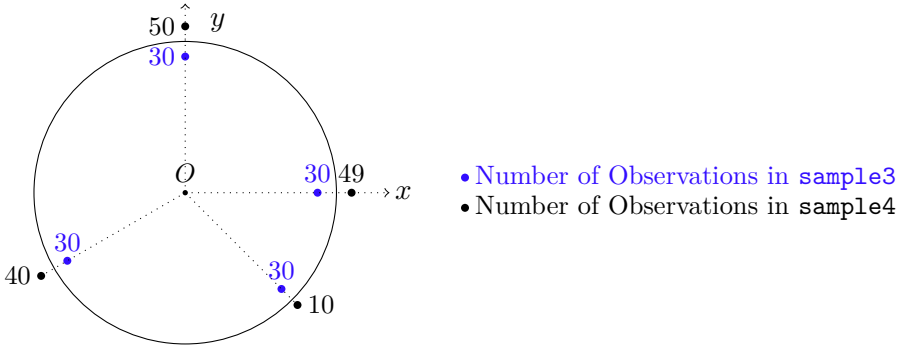


Figure 2.4: The position of observations in `sample3` and `sample4`.

The `sample3` and `sample4` contain the observations 0, $\frac{\pi}{2}$, $\frac{7\pi}{6}$, $\frac{7\pi}{4}$, which are measured from the axis Ox anticlockwise, with multiplicities 30, 30, 30, 30 for `sample3` and 49, 50, 40, 10 for `sample4`. To test whether the circular data from `sample3` and `sample4` come from the same distribution, one needs to use the function `Kuiper2sample`. Hence, the following code should be used:

```
R> sample3 <- c(rep(0,30),rep(pi/2,30),rep(7*pi/6,30),rep(7*pi/4,30))
R> sample4 <- c(rep(0,49),rep(pi/2,50),rep(7*pi/6,40),rep(7*pi/4,10))
R> Kuiper2sample(x = sample, y = sample4)
```

```
        Two-sample Kuiper Test With Ties

data:  sample3 and sample4
v = 0.18289, p-value = 0.00611
alternative hypothesis: two-sided
```

Note that the result from the Kuiper test does not depend on the choice of coordinate system. For example, if we choose a different axis, e.g., axis Oy, and record the observed angles clockwise, the result will be the same, as illustrated by the following code.

```
R> sample3b <- c(rep(0,30),rep(pi/2,30),rep(3*pi/4,30),rep(4*pi/3,30))
R> sample4b <- c(rep(0,50),rep(pi/2,49),rep(3*pi/4,10),rep(4*pi/3,40))
R> Kuiper2sample(x = sample3b, y = sample4b)

        Two-sample Kuiper Test With Ties

data:  sample3b and sample4b
v = 0.18289, p-value = 0.00611
alternative hypothesis: two-sided
```

In the next example, we provide a comparative study of the asymptotic p-value in (2.2.15) and the exact p-value when the data samples contain ties.

Example 2.13. Let us assume that we have a realization of the order statistic $\boldsymbol{Z}_{m+n}$ as follows:

$$z_1 = z_2 = \cdots = z_{1/2(m+n)} < z_{1/2(m+n)+1} < \cdots < z_{4/5(m+n)} = z_{4/5(m+n)+1} = \cdots = -z_{m+n} \tag{2.3.1}$$

Such a sample $\boldsymbol{Z}_{m+n}$ may come from, e.g. an excess of loss reinsurance contract, with retention level M and limiting level L, see e.g. Example 3.1 in Dimitrova et al. (2020). In order to apply (2.2.15) to compute the asymptotic p-value, one has to extract the jump structure of E from (2.3.1), which yields $\Lambda = 2$, $\hat{f}_0 = \hat{f}_1 = 0$, $\hat{f}_2 = 0.5$, $\hat{f}_3 = 0.8$, $\hat{f}_4 = \hat{f}_5 = 1$, $\omega_1 = \omega_2 = 1$, $\nu_1 = 1$ and $v_0 = 0$, $v_1 = 0$, $v_2 = 1$, $w_0 = 0$, $w_1 = 1$, $w_2 = 1$. Note that the change of the sample sizes m and n does not affect the

inferred jump structure of E. According to (2.2.15), the corresponding asymptotic distribution of the KS test is

$$\hat{\Phi}(x) = [\prod_{l=1}^{3}(f_{l+1} - f_l)]^{-\frac{1}{2}} \sum_{j=-\infty}^{\infty} \frac{(-1)^j}{2\pi} \int_{-x}^{x}\int_{-x}^{x} e^{\psi} dz_1 z_2, \tag{2.3.2}$$

where

$$\psi = -\frac{1}{2}[\frac{z_1^2}{\hat{f}_2 - \hat{f}_1} + \frac{z_2^2}{\hat{f}_4 - \hat{f}_3} + \frac{z_2 - (-1)^j z_1 - 2jx}{\hat{f}_3 - \hat{f}_2}].$$

To compare the asymptotic p-value with the exact p-value, we need to implement $1 - \hat{\Phi}(x)$ in (2.3.2) and compare with $\mathbb{P}\{D_{m,n} \geq x\sqrt{\frac{n+m}{nm}}\}$ for different combinations of m and n. In this example, these p-values are computed using Mathematica and R function `KS2sample` respectively, for $x = 1$. The results are presented in Table 2.2. It is clear that when $m+n$ increases, regardless of the value of η, the exact p-values numerically converge to the asymptotic p-value. Their relative error is less than 1% when $m + n \geq 40,000$. Comparing the run times and accuracy, we see that using the asymptotic formula leads to significant efficiency gains already for $m + n \geq 90,000$.

2.4 Comparison with Existing Statistical Software

The purpose of this section is to compare the numerical performance of the methods for computing KS and Kuiper p-values presented in Section 2.2 and implemented in the R functions `KS2sample` and `Kuiper2sample`. In Sections 2.4.1 and 2.4.2, we first summarize the existing implementations of the two-sample KS and Kuiper tests in the mainstream statistical software. We discuss their functionalities and select some representative implementations of the two-sample KS and Kuiper tests which we then compare with our R functions. In order to assess the accuracy, we use our Mathematica implementations of KS and Kuiper tests to compute the true p-values and estimate the relative error for all competing implementations. For speed comparison, we also provide the corresponding CPU run times on a machine with a 3.00GHz Core i7-9700 processor with 32GB RAM, running Windows 10.

2.4.1 Comparison of Different Implementations for the KS Test

This section summarizes the existing software and packages that implement the two-sample KS test and demonstrates the advantages of our R function `KS2sample`.

η	$m+n$	q	Exact		Asympt.	Rel.err(%)
0.2	2,500	0.05	0.171417114	(0.00)	0.174525238	1.78
	10,000	0.025	0.172856075	(0.01)	(1)	0.96
	40,000	0.0125	0.173661725	(0.17)		0.49
	90,000	0.008333	0.173943119	(0.90)		0.33
	250,000	0.005	0.174172870	(7.13)		0.20
	1,000,000	0.0025	0.174347892	(114)		0.10
0.3	2,500	0.043644	0.167449717	(0.00)	0.174525238	4.05
	10,000	0.021822	0.171025013	(0.02)	(1)	2.01
	40,000	0.010911	0.172937728	(0.24)		0.91
	90,000	0.007274	0.173593027	(1.20)		0.53
	250,000	0.004364	0.173813623	(9.49)		0.41
	1,000,000	0.002182	0.174214245	(266)		0.18
0.4	2,500	0.040825	0.168216436	(0.00)	0.174525238	3.61
	10,000	0.020412	0.172763135	(0.02)	(1)	1.01
	40,000	0.010206	0.173546483	(0.27)		0.56
	90,000	0.006804	0.173817964	(1.36)		0.41
	250,000	0.004082	0.174038893	(11.1)		0.28
	1,000,000	0.002041	0.174206779	(186)		0.18

Table 2.2: Exact p-values $\mathbb{P}\{D_{m,n} \geq q\}$ when $q = \sqrt{\frac{n+m}{nm}}$ and $\frac{m}{m+n} = \eta$ obtained from `KS2sample` and asymptotic p-values $1 - \hat{\Phi}(1)$ computed by implementing (2.3.2) in Mathematica. Numbers in () are relative run times to the Mathematica implementation of (2.3.2).

To the best of our knowledge, the majority of the existing implementations of the KS test have at least one of the following limitations: ties in the sample are ignored; approximate or resampling methods rather than exact methods are used to compute p-values; no flexibility with respects to the choice of weight function is allowed. The first limitation usually results in a higher p-value when the pooled sample has ties, therefore making the test too conservative when the data samples are discrete or mixed. The second limitation results in loss of accuracy. The third limitation appears in all the alternative implementations that we have discussed, and it affects the power of the KS test as evidenced by the examples in Appendix 2.C.

First, we summarize the implementations of computing p-values for the weighted KS test. We find that the R function `ks` in the package **WRS** (Wilcox and Schönbrodt, 2022, see also Wilcox 2012) provides an exact p-value for the KS test with the Anderson-Darling weight, as defined in (2.2.1). Other than that, the R package **Ecume** (de Bezieux, 2024) implements a different weighted KS test, i.e., gives weights on the tested samples separately, rather than on the edf of the pooled

sample $W(E_{m+n}(x))$ as defined in (2.2.1).

Secondly, we focus on the methods to compute p-values for the unweighted KS test, used in alternative software implementations. In order to classify the latter, we introduce four descriptors taking values 1, 2, 3, 4 (as shown in Table 2.3) based on two criteria: whether ties are, or are not allowed, and whether the result is an exact or approximate p-value.

Table 2.3: Descriptors for the unweighted KS test.

	No Ties	Allow Ties
Non-exact	1 (Approximation)	3 (Resampling)
Exact	2	4

The result of this classification is as follows.

1 **HypothesisTests.jl** (Lin et al., 2019) in Julia (Bezanson et al., 2024), Excel package **Real Statistics Resource Pack** (Zaiontz, 2024, see function `KS2TEST`)

2 S-PLUS (TIBCO Software Inc., 2010, see function `ks.gof`), Python package **SciPy** (Virtanen et al., 2020, see function `scipy.stats.ks_2samp`)

1,2 SAS (SAS Institute Inc., 2023) procedure **PROC NPAR1WAY**, Stata (StataCorp, 2023, see function `ksmirnov`), package **NSM3** (Schneider et al., 2023, see function `pKolSmirn`), package **dgof** (Arnold and Emerson, 2022, see function `ks.test`), Mathematica (Wolfram Research, Inc, 2024, see function `KolmogorovSmirnovTest`)

3 R package **twosamples** (Dowd, 2023, see function `ks_test`)

1,3 SPSS (IBM Corp, 2022), R package **Ecume** (de Bezieux, 2024, see function `ks_test`)

1,2,3 Project **Apache Commons Math** (Apache Software Foundation, 2024, see package **KolmogorovSmirnovTest**) in Java (Oracle Corporation, 2024)

4 MAPLE (Waterloo Maple Inc., 2024) package **KSNstat** (Brown, 2011), StatsDirect (StatsDirect Ltd., 2024), StatXact (Cytel Inc., 2019), R package **WRS** (function `ks`)

1,2,3,4 R package **stats** (R Core Team, 2024, see functions `ks.test` or `psmirnov`)

For example, SPSS KS implementation is classified by **1,3**, meaning it provides two ways of computing p-values, one based on an asymptotic formula (hence descriptor **1**), and the other one based on resampling (hence descriptor **3**). The characteristics of the different methods that lead to their classification are discussed next.

Firstly, approximate calculations (descriptor **1**) are mainly based on the asymptotic formula proposed by Smirnov (1933) (only Stata also uses normal approximation) which assumes that samples are drawn from a continuous distribution, thus it naturally fails to calculate the correct p-value when there are ties. Even if the data samples are continuous, the calculated p-values by the approximation methods are accurate only when both sample sizes are sufficiently large, and usually have non-negligible errors compared to the true p-value when samples are small. Secondly, resampling-based implementations (descriptor **3**) compute p-values by simulating the values of KS statistics for pairs of samples that are randomly splitted from the pooled sample. Hence this method allows the pooled sample to have ties. (This does not apply to the resampling implementations in R packages **stats**, **NSM3** and **dgof**, though they provide calculation options for resampling estimations, they neither are available for data with ties nor give p-value with large biases). However, the performance of the resampling method depends on the number of bootstrapping iterations. If the latter number increases, the accuracy also increases as does the computational cost. In practice, (as we have also shown in Table 2.4), the implementations under the category of descriptor **3** are neither more accurate nor are they more efficient, compared with implementions using the exact methods, falling under the category of descriptors **2** and **4**.

Hence, the most suitable choice for computing the p-value of the unweighted KS test narrows down to implementations with descriptor **4**, i.e., the exact methods allowing data samples containing ties. These implementations rely on one of the three methods: Nikiforov (1994, **KSNstat** in MAPLE and StatsDirect), Hilton et al. (1994, StatXact) and Schröer and Trenkler (1995, R function `ks` in the package **WRS** and the function `psmirnov` in the package **stats**). We have found that Nikiforov (1994), Hilton et al. (1994), and Schröer and Trenkler (1995) have independently arrived at a similar algorithm for computing KS test p-values, based on the recurrence for-

mula (2.2.11) for finding the number of trajectories that lie in the specific subset of integer-valued grid R (see Section 2.2.2).

To summarize, in order to calculate the exact p-value of the two-sided unweighted two-sample KS test, the R functions `ks` from the package **WRS** and `psmirnov` from the package **stats**, can be applied for arbitrary data samples that may come from a continuous, discrete or mixed distribution. Furthermore, `psmirnov` also incorporates both (2.2.11) and (2.2.12) for the unweighted KS test, where (2.2.12) ensures higher accuracy. As for the weighted KS test, only the R function `ks` calculates the exact p-value, only for the Anderson-Darling weight and only for samples from a continuous distribution.

Next, we demonstrate the advantages of our R function `KS2sample` compared with the implementations listed above. For this purpose, we compare `KS2sample` with the function `ks_test` from the package **twosamples**, a representative of resampling estimation (descriptor **3**), the functions `ks` and `psmirnov`, representing exact p-value calculation (descriptor **4**). We use the data samples from Example 2.17, in order to compare the speed and accuracy of computing the required p-value, using the functions listed in Table 2.4, where Mathematica stands for our Mathematica implementation.

	Unweighted		Anderson-Darling weight	
Mathematica	(2.0317)	$\mathbf{2.39402603196511}\times10^{-7}$	(2.0530)	$\mathbf{6.49544137675074}\times10^{-7}$
`KS2sample`	(0.0084)	$\mathbf{2.39402603196511}\times10^{-7}$	(0.0144)	$\mathbf{6.49544137675074}\times10^{-7}$
`psmirnov`	(0.0612)	$\mathbf{2.39402603196511}\times10^{-7}$	(-)	-
`KS2sample.c`	(0.0032)	$\mathbf{2.39402605}184402\times10^{-7}$	(0.0053)	$\mathbf{6.49544139}141867\times10^{-7}$
`psmirnov.c`	(0.0487)	$\mathbf{2.39402604}962358\times10^{-7}$	(-)	-
`ks`	(-)	`NaN`	(-)	`NaN`
twosamples	(978.41)	1.0×10^{-6}	(-)	-

Table 2.4: The calculated p-values for unweighted and weighted KS tests in Example 2.17 using different implementations, with CPU run times shown in paratheses in seconds, for computing 10 times the p-value. Number of bootstrapping iterations $= 10^6$ in the package **twosamples**.

Recall that both `KS2sample` and `psmirnov` include two methods, a direct computation of the p-value P defined in (2.2.5), using (2.2.12) and an indirect one computing the complementary p-value $1-P$ as in (2.2.7), using (2.2.11). For the latter method, we label both functions with ".c" and give the corresponding p-value P. The same labelling is also used in Appendix 2.D.

Based on the results in Table 2.4, we can see that the unweighted p-values pro-

duced by `KS2sample` and `psmirnov` using (2.2.12), have 17 correct digits compared with the true p-value from Mathematica, and the p-values from both functions using (2.2.11), have 8 correct digits compared with the true p-value. Although it is claimed by the authors of the package **WRS** that the function `ks` is suitable for computing KS p-values with no weight or Anderson-Darling weight, it does not work for sample sizes $m+n > 2500$, as seen from Table 2.4, where $m+n = 2700$. We have further tested the function `ks`, and have found out that when $n+m < 2500$, it is thousands of times slower than `KS2sample` and hundreds of times slower than `psmirnov` in terms of CPU time. In the package **twosamples**, the calculation is very slow (10^6 times slower than `KS2sample`) and produces a p-value with a large error. For the above reasons, the function `ks` and the package **twosamples** are not included in our further comparisons.

Therefore, in the sequel, we systematically compare the run time and the accuracy of the functions `KS2sample` and `psmirnov` for the unweighted KS test and continuous samples, i.e no ties. Since both `KS2sample` and `psmirnov` implement the same recurrence formulas (2.2.11) and (2.2.12), their run times are not substantially different regardless of whether there are ties or no ties in the samples. As for m and n, we choose $m+n = 100$ and 500, to cover the case of small sample sizes, 2,500 and 10,000 for medium sample sizes and 50,000 and 100,000 for large sample sizes. We choose the ratio of the smaller sample size to the sum of sample sizes, $\eta := \frac{m}{m+n} = 0.2, 0.3, 0.4, 0.5$, to cover correspondingly the cases of severely unbalanced ($\eta = 0.2$), unbalanced ($\eta = 0.3$), relatively balanced ($\eta = 0.4$) and perfectly balanced ($\eta = 0.5$) sample sizes. For each combination of sample sizes m and n, the functions `KS2sample` and `psmirnov` are used to calculate p-values P as defined in (2.2.5), for some given values of q, with $m_i \equiv 1$, $i = 1, \ldots, m+n$. The relative errors and the CPU times (for 100 repeated evaluations) are summarized in Table 2.6, which is given in Appendix 2.D.

The R functions `KS2sample` and `psmirnov` with implementation (2.2.12) typically provide 16–17 correct digits. In contrast, `psmirnov` with implementation (2.2.11) fails to provide valid values when $n+m \geq 10000$, whereas `KS2sample` with implementation (2.2.11) still provides p-values with an acceptable relative error. However, Table 2.6 shows that the relative run time of `psmirnov` to `KS2sample`

grows as $m+n$ increases, despite the fact that the same recursive formulas (2.2.11) and (2.2.12) are implemented in both functions. This can be explained by their different computation complexity. Compared with `KS2sample`, `psmirnov` needs to compute the number B in (2.2.11) or J in (2.2.12) for all points on the set R which leads to a computation complexity $O(n^2)$, whereas `KS2sample` only computes B and J for points in S, leading to complexity $O(n^{3/2})$.

In conclusion, we show in Table 2.4 that the R function `KS2sample` is very fast and accurate for both the unweighted and weighted KS tests. The only viable alternative proves to be the function `psmirnov` in the package **stats**. In the thorough comparison of the functions `KS2sample` and `psmirnov`, we have shown that `KS2sample` provides equal or higher accuracy than `psmirnov`. However, `psmirnov` has a higher computational complexity and is 1.8–900 times slower than the proposed `KS2sample` for $m+n \leq 100,000$.

2.4.2 Comparison of Different Implementations for the Kuiper Test

The purpose of this section is to demonstrate the numerical performance of the proposed R function `Kuiper2sample` and compare it with other existing software and packages that implement the two-sample Kuiper test.

First, we summarize all the existing implementations that compute p-values of the two-sample Kuiper test. These include the Python package **Kuiper** (Archibald, 2015), the STATA package **CIRCSTAT** (Cox, 1998), the Mathematica function `KuiperTest`, the SAS procedure **PROC NPAR1WAY**, the Java software **AdvancedMiner** (Algolytics Technologies, 2022, see the function `Kuipertest` therein), the R package **Kuiper.2samp** (Ruan, 2018, see the function `kuiper.2samp` therein) and the package **twosamples** (see the function `kuiper_test` therein). Among all these software, the function `kuiper_test` from the R package **twosamples** is the only one that uses the resampling method to calculate p-values, while all the rest use the asymptotic method.

These two types of methods are similar to the asymptotic and resampling methods for computing the KS p-values, corresponding to the descriptors **1** and **3** in Section 2.4.1. The asymptotic implementation provides high accuracy only for large samples with no ties. When samples contain ties, the calculated p-value is larger than the true value, which makes the Kuiper test too conservative. As for the re-

sampling method, its accuracy and speed are only determined by the number of bootstrapping iterations and are not affected by whether the samples have ties or not. However, as we have previously shown in Table 2.4 for the KS test, the re-sampling method is slow and inaccurate, hence will not be included in our further comparisons.

We compare our R function `Kuiper2sample` with the R function `kuiper.2samp` from the package **Kuiper.2samp**. The latter is a representative of the implementations using the asymptotic method. We provide the corresponding run times and relative errors to the true p-values from our Mathematica implementation for the Kuiper test, as we have done in Section 2.4.1.

Similarly as in Section 2.4.1, we choose the sample sizes m and n such that $m+n$ = 100, 500, 2,500, 10,000, 50,000 and 100,000 and the ratio of the smaller sample size to the sum $\eta := \frac{m}{m+n} = 0.2, 0.3, 0.4, 0.5$, thus covering all the possible combinations of data balancing and sample sizes. Recall that `Kuiper2sample` includes two methods, a direct computation of the p-value P', defined in (2.2.22), using (2.2.27) and an indirect one computing the complementary p-value $1 - P'$, using (2.2.23). For the latter method, we label both functions with ".c" and summarize the corresponding relative errors and CPU times for 5 repeated evaluations of the p-values in Table 2.7, which is presented in Appendix 2.D.

In Table 2.7, we show that when $m + n \leq 100000$, `Kuiper2sample` provides significantly more accurate results than `kuiper.2samp`. More preciously, for `Kuiper2sample`, the relative error is typically less than 6×10^{-5} and rarely up to 2×10^{-4} for implementation (2.2.23), and is less than 2×10^{-8} for implementation (2.2.27). Implementation (2.2.23) is 1.3–2.6 times faster than implementation (2.2.27). The run time of both methods in `Kuiper2sample` grows as sample sizes $m+n$ increase, whereas the time of `kuiper.2samp` stays constant as it is asymptotic.

As discussed in Section 2.2.5, the first loop to calculate N' in (2.2.23) and the loop to calculate P' in (2.2.27) have lengths related to the size of the least common multiple of m and n, $r = lcm(m, n)$. In both loops, given an index i, the computation is with complexity $O(n^{3/2})$. Hence the total complexity is $O(n^{3/2}r)$. Therefore, for two pairs of samples with similar sample sizes, m close to m' and n close to n', but with very different least common multiples r and r', the running times may be

substantially different. An example of this is shown in Table 2.5. When $n = m$, i.e., the least common multiple is small ($r = m$), the execution time is low even for large samples, however when $n = m+1$, i.e., they are relatively prime ($r = mn$), the execution time is very high. When m is fixed and $n \approx m$, the execution time ratio for two cases is approximately the ratio between the two least common multiples. Thus for example, when $m = 3000$, one should expect that the run time for $n = m+1$ and that for $n = m+2$ is approximately 3000 and 1500 times slower than the speed when $n = m$ respectively. Therefore, for an efficient p-values computation with large samples, it is ideal to have a high greatest common divisor of sample sizes.

Table 2.5: p-value of Kuiper test and the CPU time to compute 100 times using `Kuiper2sample` when $n = m$, $m+1$ and $m+2$ and the observed value of statistic $q = 1.5\sqrt{\frac{m+n}{mn}}$.

	CPU time			Calculated p.value		
m	$n = m$	$n = m+1$	$n = m+2$	$n = m$	$n = m+1$	$n = m+2$
30	0.00	0.07	0.04	0.102592	0.083416	0.105218
50	0.00	0.27	0.15	0.084878	0.115417	0.112476
150	0.04	5.32	2.70	0.155176	0.137201	0.137160
300	0.17	50.77	25.56	0.154374	0.142191	0.148842
500	0.42	207.11	103.71	0.152341	0.155741	0.154288
1500	3.12	4624.70	2328.81	0.158969	0.165930	0.163623
2000	5.28	10647.56	5359.13	0.169913	0.164047	0.165726
3000	11.55	34454.99	17353.53	0.164606	0.169488	0.167627

However in reality, a restriction on sample sizes is impractical. Thus for a user, an alternative solution to avoid heavy computations for arbitrary m and n is to approximate p-value by calculating p-values for close sample sizes of $m' \approx m, n' \approx n$ but with large greatest common divisor. The validation of this approximation is shown in Table 2.5, where p-values for $n = m$, $m+1$ and $m+2$ are compared. The differences between these p-values decrease as m increases and become negligible for large sample sizes.

2.5 Conclusion

In this work, we have considered an alternative formulation of the two sample KS test, defined over the space of all possible pairs of samples, randomly drawn without replacement from the pooled sample. We term this permutation KS test and show in Theorem 2.1 that the difference between the critical values of the permutation KS

test and the KS test defied in terms of the ecdfs of the two samples is asymptotically negligible which therefore applies to the difference between their p-values. We use this result to develop a numerically efficient recurrence method for computing p-values of the two sample KS test. A similar method but for the unweighted KS test has been independently considered by Nikiforov (1994), Hilton et al. (1994), and Schröer and Trenkler (1995), but without any theoretical justification.

In this work, we have provided its thorough theretical justification, given by Theorem 2.1 its formal description (see Sections 2.2.1 and 2.2.2) and related proofs (see Appendix 2.A) that are missing from works of these authors (see e.g. Nikiforov (1994)), followed by some enlightening examples. We have generalized Nikiforov's method to accommodate arbitrary weight functions. In Sections 2.2.2 and 2.2.5, we have extended this method to compute exact p-values of the Kuiper test, for continuous, discrete or mixed observations in the samples. To the best of our knowledge, computation of exact Kuiper p-values has only been considered for the continuous case and equal sample sizes (Maag and Stephens, 1968; Hirakawa, 1973). Following Viehmann (2021), we have further enhanced the algorithm in order to improve the accuracy of computing KS and Kuiper p-values. We have implemented the two versions of the latter algorithm in the corresponding R functions `KS2sample` and `Kuiper2sample`, and also using Mathematica. We illustrate the application of these functions and the corresponding R code in Section 2.3.

In Section 2.2.3 we have derived a closed form formula (c.f. (2.2.15), Theorem 2.5) for the asymptotic distribution, of the two-sample Kolmogorov-Smirnov test, $D_{m,n}$, as $\frac{m}{m+n} \to \eta \in (0,1)$ and $m \to \infty$, which is valid for arbitrary samples, allowing ties in the observations. To the best of our knowledge, the asymptotic distribution of the two-sample KS test has not been investigated in the literature in the case of tied observations. Therefore, formula (2.2.15) and Theorem 2.5 represent a novel contribution of theoretical and numerical importance. We have demonstrated (see Example 2.13 and Table 2.2) that the asymptotic formula (2.2.15) is a numerically efficient alternative to the exact method, for computing KS p-values when $m+n \geq 90,000$.

Our next goal in this chapter has been to provide a thorough review of the properties of the major statistical software packages that compute p-values of the

two-sample KS and Kuiper tests (see Section 2.4). This review shows that among the nineteen software implementations, only the function `psmirnov` provides p-values with the same accuracy (up to 17 correct digits) as our proposed R function `KS2sample`, but at a higher computation cost (see Table 2.4 and Table 2.6, Appendix 2.D). For the Kuiper test, `Kuiper2sample` is the only function that provides exact p-values with at least 8 correct digits (see Table 2.7, Appendix 2.D).

Finally, we have also investigated the power, as a function of the ratio $\eta = \frac{m}{m+n}$, of the KS test for different weight functions, and of the Kuiper test. As can be seen from Examples 2.18 and 2.19, Appendix 2.C, selecting the appropriate weight function can significantly improve the power of the KS test for various combinations of the sample sizes m and n ($\eta = \frac{m}{m+n}$), that are often determined before the observations are collected. As can also be seen in Figure 6, the power of the Kuiper test, in the case of samples of discrete observations, is significantly higher than that of the KS test, regardless of the choice of weight function. This highlights the relevance of using the Kuiper test as an alternative to the KS test, and therefore of using the function `Kuiper2sample` to compute its p-values.

Appendix for Chapter 2

2.A Proofs for Chapter 2

In order to prove Theorem 2.1, let us first give the following lemmas.

Lemma 2.14. *Denote by μ_E the probability measure of a distribution with cdf $E(x)$. Let $\mathcal{F}_W$ be a class of functions $\mathcal{F}_W = \{W(E(t))\mathbb{1}_{(-\infty,t]}(\cdot) : t \in \mathbb{R}\}$. If W is bounded and uniformly continuous, given the samples $\boldsymbol{X}_m$ and $\boldsymbol{Y}_n$ that come from F and G respectively, we have*

$$\lim_{m,n\to\infty, \frac{m}{m+n}\to\eta} \mathbb{P}(\sqrt{\frac{nm}{n+m}} D^+_{m,n} \leq x) = \mathbb{P}(\sup_{f\in\mathcal{F}_W} \int f d\mathbb{B}_E \leq x), \tag{2.A.1}$$

where $\mathbb{B}_E$ is a tight brownian bridge with respect to the measure μ_E.

Proof. Denote by $\mu_{E_{m+n}}$ the empirical measure of the sample $\boldsymbol{Z}_{m+n}$, i.e. $\mu_{E_{m+n}} = \frac{1}{m+n}\sum_{l=1}^{m+n} \delta_{Z_l}$, where δ_x is the Dirac measure. For any set $A \subset \mathbb{R}$,

$$\delta_x(A) = \mathbb{1}_A(x).$$

Denote by $\tilde{\mu}_{F_m}$ the empirical measure of the sample $\tilde{\boldsymbol{X}}_m$, $\tilde{\mu}_{G_n}$ the empirical measure of the sample $\tilde{\boldsymbol{Y}}_n$, where $\tilde{\boldsymbol{X}}_m$ and $\tilde{\boldsymbol{Y}}_n$ are drawn at random from the pooled sample $\boldsymbol{Z}_{m,n}$. Let us note that

$$\sup_{f\in\mathcal{F}} \int f d(\tilde{\mu}_{F_m} - \tilde{\mu}_{G_n}) = D^{0+}_{m,n}, \tag{2.A.2}$$

where $\mathcal{F} = \{\mathbb{1}_{(-\infty,t]}(\cdot) : t \in \mathbb{R}\}$, and $D^{0+}_{m,n}$ is the unweighted one-sided KS test. Furthermore, the class $\mathcal{F}$ has a square integrable envelope function $\mathbb{1}_{\mathbb{R}}(\cdot)$, under the measures μ_F and μ_G, i.e. $\int \mathbb{1}^2_{\mathbb{R}} d\mu_F = \int \mathbb{1}^2_{\mathbb{R}} d\mu_G = 1 < \infty$. By applying Theorem 3.8.2 of van der Vaart and Wellner (2023), given the samples $\boldsymbol{X}_m$ and $\boldsymbol{Y}_n$ come from F

and G respectively, we have

$$\sqrt{\frac{nm}{n+m}}(\tilde{\mu}_{F_m} - \tilde{\mu}_{G_n}) \xrightarrow{w} \mathbb{B}_E,$$

on the class $\mathcal{F}$ almost surely, where $\xrightarrow{w}$ stands for weak convergence. By the Portmanteau theorem (see e.g. Theorem 1.3.4 in van der Vaart and Wellner, 2023), it follows that

$$\lim_{m,n\to\infty, \frac{m}{m+n}\to\eta} \mathbb{P}(\sqrt{\frac{nm}{n+m}} D^{0+}_{m,n} \leq x) = \mathbb{P}(\sup_{f\in\mathcal{F}} \int f d\mathbb{B}_E \leq x). \tag{2.A.3}$$

In addition, the limiting distribution in (2.A.3) is continuous with respect to x.

In a more general case

$$\sup_{f\in\mathcal{F}_W} \int f d(\tilde{\mu}_{F_m} - \tilde{\mu}_{G_n}) = \sup_{x\in\mathbb{R}} [\tilde{F}_m(x) - \tilde{G}_n(x)] W(E(x)). \tag{2.A.4}$$

Since W is bounded, i.e. $C = \sup_{0<t<1} W(t) < \infty$, the class $\mathcal{F}_W$ has a square integrable envelope function $C\mathbb{1}_{\mathbb{R}}(\cdot)$, under the measures μ_F and μ_G, i.e. $\int C^2 \mathbb{1}^2_{\mathbb{R}} d\mu_F = \int C^2 \mathbb{1}^2_{\mathbb{R}} d\mu_G = C^2 < \infty$. Again, by applying the Portmanteau theorem and Theorem 3.8.2 in van der Vaart and Wellner (2023), we have

$$\lim_{m,n\to\infty, \frac{m}{m+n}\to\eta} \mathbb{P}(\sqrt{\frac{mn}{m+n}} \sup_{f\in\mathcal{F}_W} \int f d(\tilde{\mu}_{F_m} - \tilde{\mu}_{G_n}) \leq x) = \mathbb{P}(\sup_{f\in\mathcal{F}_W} \int f d\mathbb{B}_E \leq x), \tag{2.A.5}$$

Since the probability in (2.A.3) is continuous with respect to x, the probability in (2.A.5) is also continuous when W is continuous. Furthermore, we have

$$\begin{aligned}
&\sqrt{\frac{mn}{m+n}} |\sup_{f\in\mathcal{F}_W} \int f d(\tilde{\mu}_{F_m} - \tilde{\mu}_{G_n}) - D^+_{m,n}| \\
\leq &\sqrt{\frac{mn}{m+n}} \sup_{x\in\mathbb{R}} \left[|\tilde{F}_m(x) - \tilde{G}_n(x)| |W(E_{m+n}(x)) - W(E(x))| \right] \\
\leq &\left[\sqrt{\frac{mn}{m+n}} \sup_{x\in\mathbb{R}} |\tilde{F}_m(x) - \tilde{G}_n(x)| \right] \sup_{x\in\mathbb{R}} |W(E_{m+n}(x)) - W(E(x))|
\end{aligned}$$

By the Glivenko–Cantelli theorem with respect to F_m and G_n, we know that $\sup_{x\in\mathbb{R}} |F_m(x) - F(x)| \to 0$ and $\sup_{x\in\mathbb{R}} |G_n(x) - G(x)| \to 0$ almost surely. Therefore when $\frac{m}{m+n} \to \eta$, $\sup_{x\in\mathbb{R}} |E_{m+n}(x) - E(x)| \to 0$ almost surely. Since W is uniformly

continuous

$$\sup_{x\in\mathbb{R}}|W(E_{m+n}(x))-W(E(x))|\to 0 \tag{2.A.6}$$

almost surely. In addition, (2.A.3) implies that $\sqrt{\frac{mn}{m+n}}\sup_{x\in\mathbb{R}}|\tilde{F}_m(x)-\tilde{G}_n(x)| = O_p(1)$, therefore

$$\sqrt{\frac{mn}{m+n}}|\sup_{f\in\mathcal{F}_W}\int f d(\tilde{\mu}_{F_m}-\tilde{\mu}_{G_n})-D^+_{m,n}|\xrightarrow{\mathbb{P}}0,$$

as $m,n\to\infty$ and $\frac{m}{m+n}\to\eta$. By the Portmanteau theorem and the continuity in (2.A.5) with respect to x, we also have that

$$\mathbb{P}(\sqrt{\frac{mn}{m+n}}\sup_{f\in\mathcal{F}_W}\int f d(\tilde{\mu}_{F_m}-\tilde{\mu}_{G_n})\le x)-\mathbb{P}(\sqrt{\frac{mn}{m+n}}D^+_{m,n}\le x)\xrightarrow{\mathbb{P}}0 \tag{2.A.7}$$

The proof is completed by combining (2.A.5) and (2.A.7). □

Lemma 2.15. *When both $\boldsymbol{X}_m$ and $\boldsymbol{Y}_n$ come from the pooled distribution E, if W is bounded and uniformly continuous*

$$\lim_{m,n\to\infty,\frac{m}{m+n}\to\eta}\mathbb{P}(\sqrt{\frac{nm}{n+m}}\Delta^+_{m,n}\le x)=\mathbb{P}(\sup_{f\in\mathcal{F}_W}\int f d\mathbb{B}_E\le x), \tag{2.A.8}$$

Proof. It is well known that

$$\sqrt{m}[\mu_{F_m}-\mu_E]\xrightarrow{w}\mathbb{B}_E \tag{2.A.9}$$

on the class $\mathcal{F}=\{\mathbb{1}_{(-\infty,t]}(\cdot):t\in\mathbb{R}\}$, where μ_{F_m} denotes the empirical measure of $\boldsymbol{X}_m$. Due to the independence of $\boldsymbol{X}_m$ and $\boldsymbol{Y}_n$, we have

$$\sqrt{\frac{mn}{m+n}}[\mu_{F_m}-\mu_{G_n}]\xrightarrow{w}\sqrt{1-\eta}\mathbb{B}_E+\sqrt{\eta}\mathbb{B}'_E \tag{2.A.10}$$

where $\mathbb{B}_E$ and $\mathbb{B}'_E$ are two independent tight Brownian bridges with respect to the measure E. Hence

$$\sqrt{\frac{mn}{m+n}}\Delta^{0+}_{m,n}=\sqrt{\frac{mn}{m+n}}\sup_{f\in\mathcal{F}}\int f d(\mu_{F_m}-\mu_{G_n})\xrightarrow{d}\sup_{f\in\mathcal{F}}\int f d(\sqrt{1-\eta}\mathbb{B}_E+\sqrt{\eta}\mathbb{B}'_E).$$

Combining the independence of $\mathbb{B}_E$ and $\mathbb{B}'_E$, we have

$$\lim_{m,n\to\infty,\frac{m}{m+n}\to\eta} \mathbb{P}(\sqrt{\frac{nm}{n+m}}\Delta^{0+}_{m,n} \le x) = \mathbb{P}(\sup_{f\in\mathcal{F}} \int f d\mathbb{B}_E \le x).$$

Similarly as in the proof of Lemma 2.14, (2.A.5) and (2.A.7) also holds when substituting $\tilde{\mu}_{F_m} - \tilde{\mu}_{G_n}$ and $D^+_{m,n}$ with $\mu_{F_m} - \mu_{G_n}$ and $\Delta^+_{m,n}$. Thus (2.A.8) holds. □

Proof of Theorem 2.1. Due to the Portmanteau theorem and the continuity in (2.A.5), one only needs to show that

$$\lim_{m,n\to\infty,\frac{m}{m+n}\to\eta} \mathbb{P}(\sqrt{\frac{nm}{n+m}}\Delta_{m,n} \le x) = \mathbb{P}(\sup_{f\in\mathcal{F}_W} |\int f d\mathbb{B}_E| \le x)$$

when both $\boldsymbol{X}_m$ and $\boldsymbol{Y}_n$ are from the same distribution E and the conditional probability

$$\lim_{m,n\to\infty,\frac{m}{m+n}\to\eta} \mathbb{P}(\sqrt{\frac{nm}{n+m}}D_{m,n} \le x) = \mathbb{P}(\sup_{f\in\mathcal{F}_W} |\int f d\mathbb{B}_E| \le x)$$

given that $\boldsymbol{X}_m$ and $\boldsymbol{Y}_n$ come from the F and G respectively. Since

$$\Delta_{m,n} = \max\{\Delta^+_{m,n}, \Delta^-_{m,n}\} \text{ and } D_{m,n} = \max\{D^+_{m,n}, D^-_{m,n}\},$$

the proof is completed by combining Lemmas 2.14 and 2.15 and the continuous mapping theorem. □

In order to prove Proposition 2.2, let us first introduce the following lemma.

Lemma 2.16. *For a pair of edfs $F_m(x,\omega)$ and $G_n(x,\omega)$, $\omega\in\Omega$ and their corresponding trajectory $\{(i_l,j_l)\}_{l=0}^{m+n}$, the value of the KS statistic $D_{m,n} = \sup_x \langle F_m(x) - G_n(x)\rangle W(E_{m+n}(x))$, calculated based on the pair $F_m(x,\omega)$ and $G_n(x,\omega)$ can be alternatively calculated as*

$$D_{m,n} = \max_{1\le l\le k} d(i_{T_l}, j_{T_l}), \tag{2.A.11}$$

where $d(i,j) = \langle i/m - j/n\rangle W((i+j)/(m+n))$, and the bracket expression $\langle\cdot\rangle$ takes either $|\cdot|$, $(\cdot)$ or $-(\cdot)$ depending on the definition of the KS test in (2.2.1) and Table 2.1 in Section 2.2.2.

Proof. In the ordered, pooled sample $(z_1 \le z_2 \le \ldots \le z_{m+n})$, according to the definition of m_l and T_l, $z_{T_l} < z_{1+T_l}$ is always true when $l < k$. Hence, $(z_{T_l} \ge z_i) \Longrightarrow (T_l \ge i)$

for $1 \leq l \leq k$ (it is also true for $l = k$ since $T_k = m+n$). Therefore, the definition of a trajectory in (2.2.8) implies that $i_{T_l} = \sum_{t=1}^{m} \mathbb{1}_{X_t \leq x}(z_{T_l})$ and $j_{T_l} = \sum_{t=1}^{n} \mathbb{1}_{Y_t \leq x}(z_{T_l})$.

Therefore, for $T_{l-1} < r \leq T_l$ $(1 \leq l \leq k)$, $z_r = z_{T_l}$, the following condition holds:

$$i_{T_l} = \sum_{t=1}^{m} \mathbb{1}_{X_t \leq x}(z_r) \quad j_{T_l} = \sum_{t=1}^{n} \mathbb{1}_{Y_t \leq x}(z_r)$$

Following the definition of F_m and G_n, we have $i_{T_l}/m = F_m(z_r)$, $j_{T_l}/m = G_n(z_r)$ and $(i_{T_l} + j_{T_l})/(m+n) = E_{m+n}(z_r)$. Hence, these relations imply that

$$d(i_{T_l}, j_{T_l}) = D_{m,n}(z_r) \quad T_{l-1} < r \leq T_l, \quad 1 \leq l \leq k, \tag{2.A.12}$$

where $D_{m,n}(x) = \langle F_m(x) - G_n(x) \rangle W(E_{m+n}(x))$. We further notice that $D_{m,n}(x)$ is a step function with all its jumps occurring at values of z_r $(1 \leq r \leq m+n)$. Hence the statement of the lemma now follows by taking the maximum on both sides of (2.A.12), with respect to r $(1 \leq r \leq m+n)$. □

Equality, (2.A.12) in Lemma 2.16 explains the meaning of $d(i_l, j_l)$ for a fixed l, which is the difference of two edfs at the point Z_l. Furthermore, it shows that the value of the KS statistic $D_{m,n}$ corresponding to a trajectory only depends on the $d(i,j)$ for $i+j = T_1, \ldots, T_k$, rather than on all the values of $d(i,j)$. Now we are in a position to prove Proposition 2.2. The core of the proof is to show that the constraint on a trajectory to lie wholly in S is equivalent to the constraint on the value of KS statistic $D_{m,n} < q$, for the pair of samples to which the trajectory corresponds.

Proof of Proposition 2.2. From definition (2.2.8) it follows that, there is a one-to-one correspondence between a trajectory and its related pair $F_m(x,\omega)$ and $G_n(x,\omega)$. Hence, to prove the proposition, we only need to prove its equivalent statement:

For a trajectory $\{(i_l, j_l)\}_{l=0}^{m+n}$, the corresponding KS statistic $D_{m,n} < q \iff \{(i_l, j_l)\}_{l=0}^{m+n} \subset S$.

Sufficiency:Assume $\{(i_l, j_l)\}_{l=0}^{m+n} \subset S$, then from the definition of S, $d(i_{T_k}, j_{T_k}) < q$ for $1 \leq k \leq K$. Based on Lemma 2.16, the latter implies $D_{m,n} < q$.

Necessity: (Proof by contradiction) Let us assume that there exists a trajectory with $D_{m,n} < q$ and $(i_{l_0}, j_{l_0}) \notin S$ for some l_0. If $l_0 = T_{k_0}$, for some $1 \leq k_0 \leq K \implies$

$d(i_{T_k}, j_{T_k}) \geq q$, leading to a contradiction to Lemma 2.16. Therefore, $l_0 \neq T_k$, for any $1 \leq k \leq K$. Then there must exist $1 \leq k_1 \leq K$, such that $T_{k_1-1} < l_0 < T_{k_1}$. Without loss of generality, let us assume either $i_{l_0} < I_{min} = \min\{i : d(i,j) < q, i+j = T_{k-1}\}$ or $i_{l_0} > I_{max} = \max\{i : d(i,j) < q, i+j = T_k\}$ is true. If the former is true, $i_{T_{k_1-1}} \leq i_{l_0} < I_{min} \implies d(i_{T_{k_1-1}}, j_{T_{k_1-1}}) \geq q$, and if the latter is true, $i_{T_k} \geq i_{l_0} > I_{max}$ $\implies d(i_{T_{k_1}}, j_{T_{k_1}}) \geq q$, both leading to a contradiction. The same logic applies to j_{l_0}, i.e., either $j < \min\{j : d(i,j) < q, i+j = T_{k-1}\}$ or $j > \max\{j : d(i,j) < q, i+j = T_k\}$ leading again to a contradiction. □

Proof of Theorem 2.5. On one hand, from (2.A.9), we have

$$\mathbb{P}(\sup_{f\in\mathcal{F}} |\int f d\mathbb{B}_E| \leq x) = \lim_{m\to\infty} \mathbb{P}(\sqrt{m}\sup_t |F_m(t) - E(t)| \leq x). \tag{2.A.13}$$

when $\boldsymbol{X}_m$ comes from the distribution E. On the other hand, by (2.A.3) and continuous mapping theorem

$$\lim_{m,n\to\infty, \frac{m}{m+n}\to\eta} \mathbb{P}(\sqrt{\frac{nm}{n+m}} D^0_{m,n} \leq x) = \mathbb{P}(\sup_{f\in\mathcal{F}} |\int f d\mathbb{B}_E| \leq x). \tag{2.A.14}$$

given $\boldsymbol{X}_m$ and $\boldsymbol{Y}_n$ coming from the distribution F and G respectively. The proof is completed by combining (2.A.13), (2.A.14) and Propositions 2.3 and 2.4 in Dimitrova et al. (2020). □

Proof of Theorem 2.8. Due to the Portmanteau theorem and the continuity in (2.A.5), we only need to show that for a fixed x, when $m,n \to \infty, \frac{m}{m+n} \to \eta$, the probability

$$\mathbb{P}(\sqrt{\frac{nm}{n+m}} \varsigma_{m,n} \leq x)$$

when both $\boldsymbol{X}_m$ and $\boldsymbol{Y}_n$ are from the same distribution E would converge to the same limit as

$$\mathbb{P}(\sqrt{\frac{nm}{n+m}} V_{m,n} \leq x)$$

given $\boldsymbol{X}_m$ and $\boldsymbol{Y}_n$ coming from the F and G respectively. Since $V_{m,n} = D^{0+}_{m,n} + D^{0-}_{m,n}$ and (2.2.21), this can be shown by combining Lemmas 2.14 and 2.15 and the continuous mapping theorem. Therefore, the statement follows. □

Proof of Proposition 2.9. Equality (2.2.21) can be rewritten as $V_{m,n} = D^{0+}_{m,n} + D^{0-}_{m,n}$

in the space Ω. Denote by $N_{eq}(a,b)$, the total number of pairs $F_m(x,\omega)$ and $G_n(x,\omega)$ for which $D^{0+}_{m,n} = a$ and $D^{0-}_{m,n} = b$. Notice that $D^{0+}_{m,n}$ and $D^{0-}_{m,n}$ can only take integer values multiples of $1/r$, therefore for natural numbers $i, j \leq 2r$, we have

$$\begin{aligned} N' &= \sum_{i+j<qr} N_{eq}(\frac{i}{r}, \frac{j}{r}) \\ N_{eq}(\frac{i}{r}, \frac{j}{r}) &= N(\frac{i+1}{r}, \frac{j+1}{r}) - N(\frac{i+1}{r}, \frac{j}{r}) - N(\frac{i}{r}, \frac{j+1}{r}) + N(\frac{i}{r}, \frac{j}{r}) \\ N_{eq}(\frac{i}{r}, 0) &= N_{eq}(0, \frac{j}{r}) = 0 \end{aligned} \tag{2.A.15}$$

Equation (2.A.15) then yields (2.2.23). The proof of (2.2.25) can be obtained by similarly as proving Proposition 2.2. □

2.B Examples on the Use of R coding for Computing the p-value of the KS Test

This section provides the related code to illustrate the use of R functions `KS2sample` and `Kuiper2sample`.

Example 2.17. Assume we are given two samples of sizes, $m = 1,200$, $n = 1,500$, each having the same four distinct observations 1, 2, 3 and 4. The table shows the

sample-1	300	300	300	300
sample-2	500	400	100	500

number of times each distinct observation is repeated in each of the samples, thus there are 4 distinct observations repeated correspondingly m_i times in the pooled sample, i.e., $m_1 = 800$, $m_2 = 700$, $m_3 = 400$ and $m_4 = 800$.

To calculate the p-value of the unweighted two-sided two-sample KS test based on `sample1` and `sample2`, one can use the function `KS2sample` as follows:

```
R> sample1 <- c(rep(1,300), rep(2,300), rep(3,300), rep(4,300))
R> sample2 <- c(rep(1,500), rep(2,400), rep(3,100), rep(4,500))
R> KS2sample(x = sample1, y = sample2)

 Unweighted Two-sample Kolmogorov-Smirnov Test With Ties

data:  sample1 and sample2
```

```
d = 0.1, p-value = 2.394e-07
alternative hypothesis: two-sided
```

For these two samples from Example 2.17, one could also use `Kuiper2sample` to compute the observed Kuiper test statistic, $V_{m,n} =$ `v` and its p-value.

```
R> Kuiper2sample(x = sample1, y = sample2)

        Two-sample Kuiper Test With Ties

data:  sample1 and sample2
v = 0.18333, p-value < 2.2e-16
alternative hypothesis: two-sided
```

To use the one-sided weighted KS $\Delta^{+}_{m,n}$ with Anderson-Darling weight and test against the alternative that the unknown underlying distribution of `sample1`, $F(x)$, is greater than that of `sample2`, $G(x)$, for at least one x, one needs to specify the weight and the alternative, as follows.

```
R> KS2sample(x = sample1, y = sample2, alternative = "greater",
R>+   weight = 0.5)

 Weighted Two-sample Kolmogorov-Smirnov Test (v=0.5) With
 Ties

data:  sample1 and sample2
d = 0.1825, p-value = 3.41e-06
alternative hypothesis: greater
```

Alternatively, the p-value can also be obtained by the function `KS2sample_Rcpp` with an appropriate specification of `w_vec` and the observed value of the KS statistic, $D_{m,n} =$ `d` obtained above.

```
R> Emn <- 1:(2700-1)/2700
R> w_vec2 <- 1/sqrt(Emn*(1-Emn))
R> d <- 0.182499098772108
```

```
R> KS2sample_Rcpp(1200,1500,kind = 2, M = M1,
+                 q = d, w_vec = w_vec2, tol = tol1)

[1] 3.409892e-06
```

To test whether `sample1` and `sample2` are from the same distribution using the weighted KS statistic with weight function $W(t) = 1/[t(2-t)]^{1/2}$ (suggested in Büning, 2001), one first needs to define the weight function in R and assign it to the argument `weight` in `KS2sample`, as follows:

```
R> f <- function(t) 1 / sqrt( t * (2 - t) )
R> KS2sample(x = sample1, y = sample2, weight = f)

 User Defined Weighted Two-sample Kolmogorov-Smirnov Test With Ties

data:  sample1 and sample2
d = 0.11729, p-value = 2.503e-06
alternative hypothesis: two-sided
```

The above code illustrates the use of `KS2sample` on the samples `sample1` and `sample2` from Example 2.17 but with different weight specification. In 2.C, we further show that the choice of weight function may affect and in some cases improve the power of the KS test, which is desirable in practice. For further discussion on the choice and role of the weight function, see e.g., Büning (2001), Finner and Gontscharuk (2018).

Next, we illustrate the use of `Kuiper2sample` to compute the observed Kuiper test statistic, $V_{m,n} =$ `v` and its p-value based on the samples from Example 2.17.

```
R> Kuiper2sample(x = sample1, y = sample2)

        Two-sample Kuiper Test With Ties

data:  sample1 and sample2
v = 0.18333, p-value < 2.2e-16
alternative hypothesis: two-sided
```

2.C Power Comparisons for the KS and Kuiper Tests

Here we provide a brief comparative study on the statistical power of the KS and Kuiper tests.

The statistical power of a test is the probability of correctly rejecting the false null hypothesis when the alternative hypothesis is specified. In order to show how the weight function affects the KS test, we choose KS statistic with three different weights, the unweighted $D^0_{m,n}$, the Anderson-Darling weighted $D^{1/2}_{m,n}$ and the KS statistic with $W(t) = 1/[t(2-t)]^{1/2}$ (suggested by Büning, 2001, see also Section 2.2). The first two statistics are widely known and applied in the literature, and the last is relatively less known. Then we design two cases under which the power of these three KS test statistics and the Kuiper test statistic are compared.

To avoid changes in the weight $W(E_{m+n}(x))$ (c.f. (2.2.1)) as sample sizes change, we fix the sum of sample sizes $m+n = 500$ and change the value of $\eta = \frac{m}{m+n}$. The power of the statistic can be expressed as a function of η, which only focuses on the effect of changes in the sample sizes ratio (see in Finner and Gontscharuk, 2018). For further investigation of the effect of changes in the sample sizes, one could fix the value of η and express the power as a function of $m+n$.

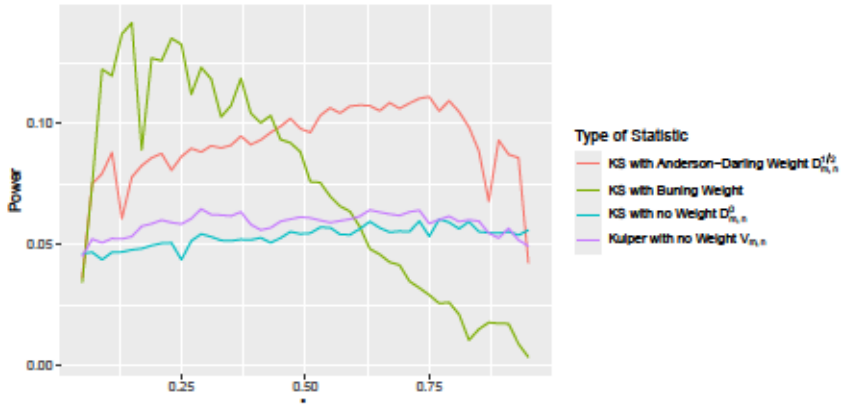


Figure 2.5: Power comparisons for samples from GEV distributions when $m+n = 500$ of the (weighted) KS tests and Kuiper test.

Example 2.18. Let the two samples come from two Generalized Extreme Value (GEV) distributions with different parameters. More preciously, the first sample of size m is drawn from a GEV distribution with cumulative distribution function $F_1(x) = \exp(-(1+0.3x)^{10/3})\mathbb{1}_{\{0.3x>-1\}}(x)$, and the second sample of size n is drawn from $F_2(x) = \exp(-(1+0.4x)^{5/2})\mathbb{1}_{\{0.4x>-1\}}(x)$, where $\mathbb{1}_A(x)$ denotes the indicator

function, which takes value 1 if $x \in A$ and takes value 0 otherwise. In order to estimate the power, for each value of η, calculations are repeated 10^5 times. As shown in Figure 2.5, the power of $D_{m,n}^{1/2}$ is higher than the power of $D_{m,n}^{0}$ across all values of η. The KS statistic with weight function suggested by Büning has the highest power when $\eta < 0.4$, and has lower power when $\eta > 0.6$.

Example 2.19. Let the two samples come from two negative binomial distributions. More precisely, the first sample is drawn from NB(6,0.75), and the second sample is drawn from NB(18,0.9). Hence, these two distributions have the same mean. The power is estimated and shown in Figure 2.6. In this case, the KS test with Büning weight has higher power than the unweighted KS statistic across all values of η, whereas $D_{m,n}^{1/2}$ has the highest power only for $0.1 < \eta < 0.5$. Moreover, we show that the Kuiper test has power higher than the KS statistic with Büning weight for all η, and that the power of these two tests are higher than that of the unweighted KS test, for all η.

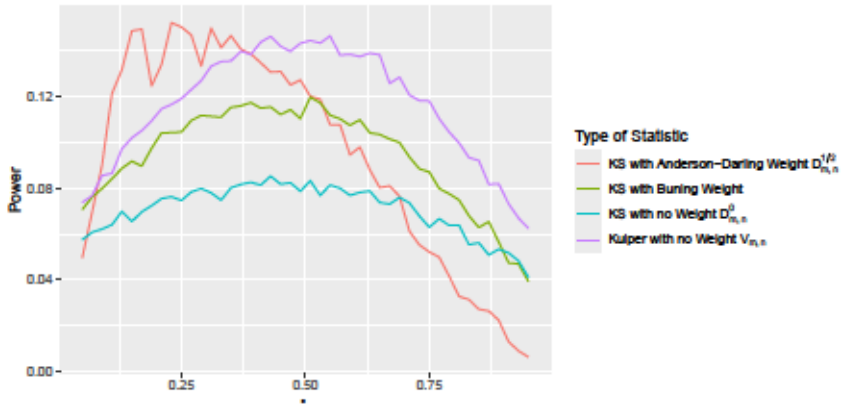


Figure 2.6: Power comparisons for samples from Negative Binomial Distributions of the (weighted) KS tests and Kuiper test when $m+n$=500.

Examples 2.18 and 2.19 demonstrate the importance of allowing for more flexibility in choosing the weight function, since as demonstrated, the KS statistic with different weights may have different power performance. Implementing the Kuiper test is also important since, as demonstrated its power is often superior to the KS test.

2.D Comparisons for the different implementations of KS and Kuiper Tests

Table 2.6: Relative error (use Mathematica as benchmark) and the CPU time for 100 times repeated evaluations (parentheses shows the relative time) of functions `KS2sample` and the Exact method of `psmirnov` for computing $\mathbb{P}(D_{m,n} \geq q)$ when $\lambda = q\sqrt{\frac{nm}{n+m}} = 3,\ 2,\ 1,\ 0.5$ for continuous underlying distributions.

λ	$m+n$		$\eta = 0.2$		$\eta = 0.3$		$\eta = 0.4$		$\eta = 0.5$	
			Rel.err(%)	Time	Rel.err(%)	Time	Rel.err(%)	Time	Rel.err(%)	Time
3	100	KS2sample	4.2×10^{-13}	1.6×10^{-3}(1.6)	2.4×10^{-13}	1.6×10^{-3}(1.45)	0	1.3×10^{-3}(1.44)	0	1.1×10^{-3}(1.22)
		psmirnov	4.2×10^{-13}	1.9×10^{-3}(1.9)	2.4×10^{-13}	2×10^{-3}(1.82)	0	2.1×10^{-3}(2.33)	0	2.1×10^{-3}(2.33)
		KS2sample.c	4×10^{-5}	1×10^{-3}(1)	8.5×10^{-6}	1.1×10^{-3}(1)	1.6×10^{-5}	9×10^{-4}(1)	3.9×10^{-6}	9×10^{-4}(1)
		psmirnov.c	2.6×10^{-5}	1.8×10^{-3}(1.8)	5.7×10^{-6}	2.1×10^{-3}(1.91)	7.4×10^{-6}	1.8×10^{-3}(2)	-3.4×10^{-6}	2×10^{-3}(2.22)
	500	KS2sample	0	4.6×10^{-3}(1.92)	5.3×10^{-13}	4.8×10^{-3}(1.78)	0	5×10^{-3}(1.85)	6.4×10^{-13}	5.2×10^{-3}(1.86)
		psmirnov	0	0.0112(4.67)	5.3×10^{-13}	0.0138(5.11)	0	0.0155(5.74)	6.4×10^{-13}	0.016(5.71)
		KS2sample.c	8.6×10^{-6}	2.4×10^{-3}(1)	-7.3×10^{-7}	2.7×10^{-3}(1)	5.6×10^{-6}	2.7×10^{-3}(1)	8×10^{-6}	2.8×10^{-3}(1)
		psmirnov.c	4.1×10^{-6}	0.01(4.17)	-1.3×10^{-7}	0.0123(4.56)	-2.1×10^{-6}	0.014(5.19)	5.8×10^{-6}	0.0147(5.25)
	2500	KS2sample	0	0.0375(2.03)	4×10^{-13}	0.045(2.24)	3.8×10^{-13}	0.045(2.22)	0	0.0461(2.23)
		psmirnov	0	0.198(10.7)	4×10^{-13}	0.264(13.1)	3.8×10^{-13}	0.304(15)	0	0.301(14.5)
		KS2sample.c	1.1×10^{-5}	0.0185(1)	1.9×10^{-5}	0.0201(1)	1.3×10^{-5}	0.0203(1)	-4.5×10^{-7}	0.0207(1)
		psmirnov.c	7.2×10^{-6}	0.218(11.8)	8.7×10^{-6}	0.279(13.9)	9.2×10^{-6}	0.322(15.8)	6×10^{-6}	0.333(16.1)
	10000	KS2sample	3.3×10^{-13}	0.279(2.4)	0	0.314(2.43)	0	0.341(2.43)	0	0.333(2.54)
		psmirnov	3.3×10^{-13}	2.96(25.5)	3.6×10^{-13}	3.85(29.8)	0	4.42(31.5)	0	4.5(34.4)
		KS2sample.c	1.6×10^{-5}	0.116(1)	-2.2×10^{-5}	0.129(1)	9.1×10^{-6}	0.14(1)	5.5×10^{-6}	0.131(1)
		psmirnov.c	3.5×10^{7}	3.8(32.8)	3.5×10^{9}	4.63(35.8)	3.4×10^{9}	5.18(36.9)	3.3×10^{9}	5.24(40)
	50000	KS2sample	3.6×10^{-13}	3.22(2.68)	3.6×10^{-13}	3.73(2.66)	0	4.01(3.34)	0	4(2.5)
		psmirnov	3.6×10^{-13}	76.2(63.5)	3.6×10^{-13}	100(71.6)	0	113(94.5)	0	117(73.1)
		KS2sample.c	0.253	1.2(1)	3.3×10^{-5}	1.4(1)	4.4×10^{-5}	1.2(1)	4.1×10^{-5}	1.6(1)
		psmirnov.c	3.4×10^{9}	92.2(76.8)	3.4×10^{9}	121(86.6)	3.4×10^{9}	138(115)	3.3×10^{9}	144(90.2)
	100000	KS2sample	0	9.94(2.92)	0	11.6(2.89)	3.3×10^{-13}	11.9(2.82)	0	11.5(2.88)
		psmirnov	3.3×10^{-13}	298(87.7)	0	389(97.4)	0	446(106)	-3.3×10^{-13}	464(116)
		KS2sample.c	1.1×10^{8}	3.4(1)	4.9×10^{-5}	4(1)	1.1×10^{-4}	4.2(1)	8.7×10^{-5}	4(1)
		psmirnov.c	3.4×10^{9}	366(108)	3.3×10^{9}	482(120)	3.3×10^{9}	551(131)	3.3×10^{9}	578(145)
2	100	KS2sample	0	1.1×10^{-3}(1.22)	2.7×10^{-13}	1.1×10^{-3}(1.22)	2.2×10^{-13}	1.2×10^{-3}(1.2)	0	1.1×10^{-3}(1.22)
		psmirnov	0	1.9×10^{-3}(2.11)	0	2×10^{-3}(2.22)	2.2×10^{-13}	2×10^{-3}(2)	0	2.1×10^{-3}(2.33)
		KS2sample.c	2.2×10^{-10}	9×10^{-4}(1)	1.6×10^{-10}	9×10^{-4}(1)	5.9×10^{-11}	1×10^{-3}(1)	-2.5×10^{-11}	9×10^{-4}(1)
		psmirnov.c	1.7×10^{-10}	1.9×10^{-3}(2.11)	1.6×10^{-10}	1.9×10^{-3}(2.11)	3.5×10^{-11}	1.9×10^{-3}(1.9)	1.3×10^{-11}	1.8×10^{-3}(2)
	500	KS2sample	1.8×10^{-13}	3.9×10^{-3}(1.86)	2×10^{-13}	4.2×10^{-3}(1.68)	1.8×10^{-13}	4.2×10^{-3}(1.91)	1.6×10^{-13}	4.2×10^{-3}(2)
		psmirnov	0	0.0106(5.05)	2×10^{-13}	0.0131(5.24)	1.8×10^{-13}	0.0145(6.59)	1.6×10^{-13}	0.0151(7.19)
		KS2sample.c	1.5×10^{-10}	2.1×10^{-3}(1)	2.8×10^{-10}	2.5×10^{-3}(1)	-8.6×10^{-12}	2.2×10^{-3}(1)	2.2×10^{-10}	2.1×10^{-3}(1)
		psmirnov.c	4.5×10^{-11}	0.0102(4.86)	4.3×10^{-11}	0.0128(5.12)	-8.6×10^{-12}	0.0144(6.55)	3×10^{-10}	0.015(7.14)
	2500	KS2sample	0	0.0291(1.98)	0	0.0315(2.19)	0	0.0333(2.11)	0	0.0338(2.13)
		psmirnov	0	0.191(13)	-1.6×10^{-13}	0.246(17.1)	0	0.279(17.7)	0	0.29(18.2)
		KS2sample.c	6.8×10^{-10}	0.0147(1)	1.3×10^{-9}	0.0144(1)	1.1×10^{-11}	0.0158(1)	1.2×10^{-10}	0.0159(1)
		psmirnov.c	1.8×10^{-10}	0.21(14.3)	4.2×10^{-10}	0.276(19.1)	3.9×10^{-10}	0.314(19.9)	2.4×10^{-10}	0.328(20.6)
	10000	KS2sample	1.6×10^{-13}	0.194(2.34)	1.6×10^{-13}	0.221(2.4)	1.6×10^{-13}	0.232(2.4)	1.6×10^{-13}	0.237(2.42)
		psmirnov	3.1×10^{-13}	2.81(33.9)	0	3.67(39.8)	1.6×10^{-13}	4.19(43.4)	-1.4×10^{-13}	4.39(44.8)
		KS2sample.c	1.1×10^{-9}	0.0829(1)	8×10^{-11}	0.0922(1)	1.1×10^{-11}	0.0966(1)	4.5×10^{-10}	0.098(1)
		psmirnov.c	1.5×10^{3}	3.64(43.9)	1.5×10^{5}	4.49(48.7)	1.5×10^{5}	5.1(52.7)	1.5×10^{5}	5.32(54.2)
	50000	KS2sample	0	2.17(2.17)	3.1×10^{-13}	2.5(2.08)	1.6×10^{-13}	2.67(2.23)	1.6×10^{-13}	2.69(2.24)
		psmirnov	4.7×10^{-13}	74.6(74.6)	0	98.1(81.8)	0	112(93.7)	0	117(97.4)
		KS2sample.c	1.6×10^{-9}	1(1)	6.6×10^{-10}	1.2(1)	1.3×10^{-9}	1.2(1)	7.1×10^{-10}	1.2(1)
		psmirnov.c	1.5×10^{5}	93(93)	1.5×10^{5}	123(102)	1.5×10^{5}	142(118)	1.5×10^{5}	146(122)
	100000	KS2sample	0	6.16(2.37)	1.6×10^{-13}	7.08(2.53)	0	7.49(2.68)	1.6×10^{-13}	7.48(2.49)
		psmirnov	4.7×10^{-13}	297(114)	-1.4×10^{-13}	389(139)	0	442(158)	1.6×10^{-13}	461(154)
		KS2sample.c	2×10^{-9}	2.6(1)	4.5×10^{-9}	2.8(1)	2.9×10^{-9}	2.8(1)	1.3×10^{-9}	3(1)
		psmirnov.c	1.5×10^{5}	377(145)	1.5×10^{5}	488(174)	1.5×10^{5}	549(196)	1.5×10^{5}	580(193)

λ	$m+n$		$\eta = 0.2$		$\eta = 0.3$		$\eta = 0.4$		$\eta = 0.5$	
			Rel.err(%)	Time	Rel.err(%)	Time	Rel.err(%)	Time	Rel.err(%)	Time
1	100	KS2sample	4×10^{-13}	9×10^{-4}(0.9)	0	1×10^{-3}(1.43)	4.4×10^{-13}	1.1×10^{-3}(1.22)	3.8×10^{-13}	1×10^{-3}(1.25)
		psmirnov	4×10^{-13}	1.9×10^{-3}(1.9)	0	1.9×10^{-3}(2.71)	4.4×10^{-13}	2×10^{-3}(2.22)	3.8×10^{-13}	2×10^{-3}(2.5)
		KS2sample.c	4×10^{-13}	1×10^{-3}(1)	0	7×10^{-4}(1)	4.4×10^{-13}	9×10^{-4}(1)	3.8×10^{-13}	8×10^{-4}(1)
		psmirnov.c	4×10^{-13}	1.8×10^{-3}(1.8)	0	1.8×10^{-3}(2.57)	4.4×10^{-13}	2×10^{-3}(2.22)	3.8×10^{-13}	2×10^{-3}(2.5)
	500	KS2sample	4×10^{-13}	3.1×10^{-3}(1.48)	4×10^{-13}	3.3×10^{-3}(1.65)	4×10^{-13}	3.3×10^{-3}(1.74)	4.2×10^{-13}	3.4×10^{-3}(1.36)
		psmirnov	4×10^{-13}	9.7×10^{-3}(4.62)	4×10^{-13}	0.012(6)	4×10^{-13}	0.0134(7.05)	4.2×10^{-13}	0.0142(5.68)
		KS2sample.c	4×10^{-13}	2.1×10^{-3}(1)	8×10^{-13}	2×10^{-3}(1)	0	1.9×10^{-3}(1)	8.2×10^{-13}	2.5×10^{-3}(1)
		psmirnov.c	4×10^{-13}	0.0104(4.95)	4×10^{-13}	0.0129(6.45)	4×10^{-13}	0.0147(7.74)	4.2×10^{-13}	0.0152(6.08)
	2500	KS2sample	0	0.0193(1.77)	0	0.0209(2.07)	3.8×10^{-13}	0.0216(2.14)	0	0.022(1.96)
		psmirnov	3.8×10^{-13}	0.182(16.7)	0	0.237(23.4)	3.8×10^{-13}	0.269(26.7)	0	0.28(25)
		KS2sample.c	1.1×10^{-12}	0.0109(1)	1.2×10^{-12}	0.0101(1)	3.8×10^{-13}	0.0101(1)	0	0.0112(1)
		psmirnov.c	7.5×10^{-13}	0.216(19.8)	7.8×10^{-13}	0.282(28)	7.8×10^{-13}	0.322(31.9)	7.3×10^{-13}	0.337(30.1)
	10000	KS2sample	3.8×10^{-13}	0.117(1.99)	3.8×10^{-13}	0.13(2.28)	3.8×10^{-13}	0.138(2.19)	0	0.139(2.18)
		psmirnov	3.8×10^{-13}	2.77(47.2)	0	3.63(63.5)	3.8×10^{-13}	4.11(65.2)	0	4.27(66.8)
		KS2sample.c	1.1×10^{-12}	0.0587(1)	0	0.0571(1)	1.1×10^{-12}	0.063(1)	-1.1×10^{-12}	0.0639(1)
		psmirnov.c	1.1×10^{-7}	3.51(59.7)	278	4.53(79.3)	274	5.11(81.1)	270	5.32(83.2)
	50000	KS2sample	0	1.2(2)	0	1.44(2.4)	0	1.44(2.4)	0	1.45(2.42)
		psmirnov	7.5×10^{-13}	74.7(124)	-3.8×10^{-13}	97(162)	0	110(183)	3.8×10^{-13}	114(191)
		KS2sample.c	1.9×10^{-12}	0.6(1)	3×10^{-12}	0.6(1)	1.1×10^{-12}	0.6(1)	6×10^{-12}	0.6(1)
		psmirnov.c	275	94(157)	273	123(205)	273	140(233)	273	143(239)
	100000	KS2sample	0	3.19(2.66)	3.8×10^{-13}	3.59(2.24)	3.8×10^{-13}	3.85(2.14)	3.8×10^{-13}	3.91(2.44)
		psmirnov	1.1×10^{-12}	290(242)	-7.4×10^{-13}	381(238)	3.8×10^{-13}	434(241)	3.8×10^{-13}	453(283)
		KS2sample.c	7.5×10^{-12}	1.2(1)	6×10^{-12}	1.6(1)	7.8×10^{-12}	1.8(1)	1.5×10^{-11}	1.6(1)
		psmirnov.c	273	367(306)	273	481(301)	272	559(310)	274	575(359)
0.5	100	KS2sample	0	9×10^{-4}(1.12)	0	8×10^{-4}(1)	0	7×10^{-4}(0.875)	1.1×10^{-13}	9×10^{-4}(0.9)
		psmirnov	0	1.7×10^{-3}(2.12)	0	1.9×10^{-3}(2.38)	1.1×10^{-13}	2.1×10^{-3}(2.62)	1.1×10^{-13}	1.9×10^{-3}(1.9)
		KS2sample.c	0	8×10^{-4}(1)	0	8×10^{-4}(1)	1.1×10^{-13}	8×10^{-4}(1)	1.1×10^{-13}	1×10^{-3}(1)
		psmirnov.c	0	1.7×10^{-3}(2.12)	0	1.9×10^{-3}(2.38)	1.1×10^{-13}	2.1×10^{-3}(2.62)	1.1×10^{-13}	2.1×10^{-3}(2.1)
	500	KS2sample	1.1×10^{-13}	2.2×10^{-3}(1.29)	1.1×10^{-13}	2.2×10^{-3}(0.957)	1.1×10^{-13}	2.2×10^{-3}(1.29)	0	2.2×10^{-3}(1.22)
		psmirnov	1.1×10^{-13}	9.1×10^{-3}(5.35)	0	0.0115(5)	1.1×10^{-13}	0.0128(7.53)	0	0.0133(7.39)
		KS2sample.c	1.1×10^{-13}	1.7×10^{-3}(1)	1.1×10^{-13}	2.3×10^{-3}(1)	1.1×10^{-13}	1.7×10^{-3}(1)	1.1×10^{-13}	1.8×10^{-3}(1)
		psmirnov.c	1.1×10^{-13}	0.0103(6.06)	1.1×10^{-13}	0.0131(5.7)	1.1×10^{-13}	0.0147(8.65)	1.1×10^{-13}	0.0152(8.44)
	2500	KS2sample	0	0.0141(1.81)	1.1×10^{-13}	0.0149(1.91)	1.1×10^{-13}	0.0153(1.7)	1.1×10^{-13}	0.0154(1.59)
		psmirnov	2×10^{-13}	0.178(22.9)	0	0.233(29.8)	1.1×10^{-13}	0.265(29.5)	1.1×10^{-13}	0.278(28.6)
		KS2sample.c	1.1×10^{-13}	7.8×10^{-3}(1)	1.1×10^{-13}	7.8×10^{-3}(1)	1.1×10^{-13}	9×10^{-3}(1)	1.1×10^{-13}	9.7×10^{-3}(1)
		psmirnov.c	1.1×10^{-13}	0.219(28.1)	1.1×10^{-13}	0.287(36.8)	1.1×10^{-13}	0.329(36.6)	1.1×10^{-13}	0.342(35.2)
	10000	KS2sample	1.1×10^{-13}	0.0745(1.77)	1.1×10^{-13}	0.0836(2.15)	0	0.0864(2.24)	2×10^{-13}	0.0885(2.05)
		psmirnov	3.1×10^{-13}	2.76(65.4)	-1×10^{-13}	3.62(93.1)	0	4.13(107)	0	4.29(99.5)
		KS2sample.c	1.1×10^{-13}	0.0422(1)	1.1×10^{-13}	0.0389(1)	0	0.0386(1)	1.1×10^{-13}	0.0431(1)
		psmirnov.c	1.1×10^{-13}	3.48(82.4)	4.12	4.6(118)	4.06	5.2(135)	3.74	5.43(126)
	50000	KS2sample	0	0.72(1.8)	0	0.8(4)	0	0.81(1.35)	0	0.85(2.12)
		psmirnov	7.3×10^{-13}	73.1(183)	-4.1×10^{-13}	94.8(474)	1.1×10^{-13}	109(182)	0	113(283)
		KS2sample.c	1.1×10^{-13}	0.4(1)	1.1×10^{-13}	0.2(1)	1.1×10^{-13}	0.6(1)	1.1×10^{-13}	0.4(1)
		psmirnov.c	3.88	92.4(231)	3.89	120(602)	3.83	140(233)	3.8	143(358)
	100000	KS2sample	-2.1×10^{-13}	1.84(1.84)	1.1×10^{-13}	2.03(1.69)	1.1×10^{-13}	2.12(3.53)	0	2.14(2.67)
		psmirnov	8.2×10^{-13}	292(292)	-4.2×10^{-13}	379(316)	3.1×10^{-13}	434(723)	-1×10^{-13}	447(559)
		KS2sample.c	1.1×10^{-13}	1(1)	1.1×10^{-13}	1.2(1)	2×10^{-13}	0.6(1)	2×10^{-13}	0.8(1)
		psmirnov.c	3.88	371(371)	3.84	480(400)	3.86	555(925)	3.94	575(719)

Table 2.7: Relative error (use Mathematica as benchmark) and the CPU time in seconds (parentheses shows the relative time) for 5 repeated evaluations of functions `kuiper.2samp` and `Kuiper2sample` for computing $\mathbb{P}(V_{m,n} \geq q)$ when $\lambda = q\sqrt{\frac{nm}{n+m}} = 3,\ 2,\ 1.5,\ 1$ for continuous underlying distributions.

λ	$m+n$		$\eta = 0.2$		$\eta = 0.3$		$\eta = 0.4$		$\eta = 0.5$	
			Rel.err(%)	Time	Rel.err(%)	Time	Rel.err(%)	Time	Rel.err(%)	Time
3	100	`Kuiper2sample`	1.9×10^{-6}	1.2×10^{-3}(1.77)	-2×10^{-6}	3.2×10^{-3}(1.75)	9.2×10^{-7}	1.8×10^{-3}(1.75)	-1.1×10^{-7}	7.2×10^{-4}(1.77)
		`Kuiper2sample.c`	5.5×10^{-6}	7×10^{-4}(1)	2.7×10^{-5}	1.8×10^{-3}(1)	-1.1×10^{-4}	1.1×10^{-3}(1)	-1.2×10^{-5}	4.1×10^{-4}(1)
		`kuiper.2samp`	66.8	8.1×10^{-5}(0.116)	0.845	8.8×10^{-5}(0.0485)	30.8	9.7×10^{-5}(0.0919)	-19.1	9.7×10^{-5}(0.238)
	500	`Kuiper2sample`	6×10^{-8}	0.03(2.4)	-5.6×10^{-8}	0.076(2.17)	1.5×10^{-6}	0.042(2.05)	-1.1×10^{-7}	0.018(2.12)
		`Kuiper2sample.c`	8.9×10^{-6}	0.0125(1)	7.7×10^{-5}	0.035(1)	1.2×10^{-5}	0.0205(1)	-6.4×10^{-5}	8.5×10^{-3}(1)
		`kuiper.2samp`	136	5.5×10^{-5}(4.4×10^{-3})	91.6	5.9×10^{-5}(1.7×10^{-3})	89.3	5.6×10^{-5}(2.7×10^{-3})	137	5.6×10^{-5}(6.5×10^{-3})
	2500	`Kuiper2sample`	4.9×10^{-7}	0.591(2.05)	-8.4×10^{-7}	1.53(2.1)	3.8×10^{-7}	0.874(2.14)	-9.5×10^{-7}	0.362(2.14)
		`Kuiper2sample.c`	-1.2×10^{-4}	0.287(1)	-2.4×10^{-4}	0.731(1)	1.7×10^{-5}	0.408(1)	1.3×10^{-4}	0.169(1)
		`kuiper.2samp`	24.2	5.6×10^{-5}(1.9×10^{-4})	27.7	5.6×10^{-5}(7.6×10^{-5})	26.6	5.6×10^{-5}(1.4×10^{-4})	14.7	5.6×10^{-5}(3.3×10^{-4})
	10000	`Kuiper2sample`	9.1×10^{-7}	8.57(2.12)	9.5×10^{-7}	22.4(2.15)	1.7×10^{-6}	12.7(2.18)	-1.8×10^{-6}	5.23(2.16)
		`Kuiper2sample.c`	-7.8×10^{-4}	4.05(1)	9.7×10^{-4}	10.4(1)	-6.2×10^{-4}	5.84(1)	4.2×10^{-4}	2.42(1)
		`kuiper.2samp`	10.1	5.5×10^{-5}(1.4×10^{-5})	11	5.6×10^{-5}(5.3×10^{-6})	9.54	5.6×10^{-5}(9.6×10^{-6})	6.45	5.6×10^{-5}(2.3×10^{-5})
	50000	`Kuiper2sample`	1×10^{-7}	202(2.16)	-4.5×10^{-7}	527(2.16)	-2.7×10^{-6}	303(2.2)	-2.5×10^{-7}	126(2.21)
		`Kuiper2sample.c`	5.8×10^{-3}	93.8(1)	-1.8×10^{-3}	244(1)	-5.8×10^{-3}	138(1)	3.1×10^{-3}	57.1(1)
		`kuiper.2samp`	5.06	5.5×10^{-5}(5.9×10^{-7})	5.34	5.6×10^{-5}(2.3×10^{-7})	5.46	6×10^{-5}(4.4×10^{-7})	3.62	5.6×10^{-5}(9.7×10^{-7})
	100000	`Kuiper2sample`	-2.4×10^{-8}	801(2.16)	-1.6×10^{-6}	2.1×10^{3}(2.17)	-2.7×10^{-6}	1.2×10^{3}(2.2)	1.2×10^{-7}	498(2.2)
		`Kuiper2sample.c`	-3.8×10^{-3}	370(1)	0.0124	966(1)	-0.0224	546(1)	1.5×10^{-3}	227(1)
		`kuiper.2samp`	4.01	5.6×10^{-5}(1.5×10^{-7})	3.45	5.7×10^{-5}(5.9×10^{-8})	2.86	5.6×10^{-5}(1×10^{-7})	3.03	5.6×10^{-5}(2.5×10^{-7})
2	100	`Kuiper2sample`	-5×10^{-12}	7.1×10^{-4}(1.69)	5.6×10^{-12}	1.8×10^{-3}(1.44)	-2.3×10^{-11}	1×10^{-3}(1.48)	-3.1×10^{-11}	4.1×10^{-4}(1.61)
		`Kuiper2sample.c`	-2.2×10^{-10}	4.2×10^{-4}(1)	-1.8×10^{-10}	1.2×10^{-3}(1)	4.5×10^{-10}	6.7×10^{-4}(1)	3.2×10^{-10}	2.6×10^{-4}(1)
		`kuiper.2samp`	78.6	6.8×10^{-5}(0.163)	92.4	6.7×10^{-5}(0.0538)	65	6.7×10^{-5}(0.0995)	41	6.8×10^{-5}(0.267)
	500	`Kuiper2sample`	2.7×10^{-11}	0.0175(2.33)	6.9×10^{-12}	0.0405(2.25)	-1.9×10^{-11}	0.0235(2.61)	3.8×10^{-11}	0.01(2.5)
		`Kuiper2sample.c`	-5.3×10^{-10}	7.5×10^{-3}(1)	2.7×10^{-9}	0.018(1)	-4×10^{-9}	9×10^{-3}(1)	6.1×10^{-10}	4×10^{-3}(1)
		`kuiper.2samp`	39	6.7×10^{-5}(8.9×10^{-3})	29.2	6.8×10^{-5}(3.8×10^{-3})	31.5	6.7×10^{-5}(7.5×10^{-3})	27.3	6.7×10^{-5}(0.0167)
	2500	`Kuiper2sample`	-7.3×10^{-12}	0.301(2.07)	7.9×10^{-11}	0.769(2.36)	-4.8×10^{-11}	0.431(2.41)	3×10^{-11}	0.179(2.32)
		`Kuiper2sample.c`	-2.9×10^{-9}	0.145(1)	-8.2×10^{-9}	0.327(1)	-3.2×10^{-8}	0.179(1)	-6.9×10^{-9}	0.077(1)
		`kuiper.2samp`	11.1	6.7×10^{-5}(4.6×10^{-4})	14.3	6.7×10^{-5}(2.1×10^{-4})	10.3	7×10^{-5}(3.9×10^{-4})	7.11	6.6×10^{-5}(8.6×10^{-4})
	10000	`Kuiper2sample`	-1×10^{-10}	4.06(2.05)	-1.8×10^{-10}	10.5(2.21)	-2.5×10^{-11}	5.97(2.15)	4.9×10^{-11}	2.48(2.12)
		`Kuiper2sample.c`	7.1×10^{-8}	1.98(1)	-1.2×10^{-7}	4.77(1)	6.1×10^{-8}	2.78(1)	-3.8×10^{-8}	1.17(1)
		`kuiper.2samp`	5.36	6.7×10^{-5}(3.4×10^{-5})	6.47	6.7×10^{-5}(1.4×10^{-5})	5.16	6.7×10^{-5}(2.4×10^{-5})	3.51	6.7×10^{-5}(5.7×10^{-5})
	50000	`Kuiper2sample`	-9.7×10^{-11}	93.9(2.18)	4.7×10^{-10}	243(2.16)	-5.8×10^{-12}	139(2.2)	1×10^{-10}	57.2(2.18)
		`Kuiper2sample.c`	-2×10^{-7}	43.1(1)	-1.4×10^{-7}	112(1)	9.5×10^{-7}	63.1(1)	-5.7×10^{-8}	26.2(1)
		`kuiper.2samp`	3.27	6.6×10^{-5}(1.5×10^{-6})	2.93	6.7×10^{-5}(6×10^{-7})	2.86	6.7×10^{-5}(1.1×10^{-6})	4.07	6.8×10^{-5}(2.6×10^{-6})
	100000	`Kuiper2sample`	3.9×10^{-10}	367(2.18)	-1.2×10^{-10}	941(2.16)	-1.4×10^{-10}	537(2.18)	3.3×10^{-10}	223(2.18)
		`Kuiper2sample.c`	-2.2×10^{-7}	168(1)	1.4×10^{-7}	436(1)	-7.3×10^{-7}	246(1)	2.5×10^{-6}	102(1)
		`kuiper.2samp`	1.76	6.7×10^{-5}(4×10^{-7})	2.12	7×10^{-5}(1.6×10^{-7})	2.25	6.7×10^{-5}(2.7×10^{-7})	2.32	6.8×10^{-5}(6.7×10^{-7})

λ	$m+n$		$\eta = 0.2$		$\eta = 0.3$		$\eta = 0.4$		$\eta = 0.5$	
			Rel.err(%)	Time	Rel.err(%)	Time	Rel.err(%)	Time	Rel.err(%)	Time
1.5	100	Kuiper2sample	1.6×10^{-12}	4.8×10^{-4}(1.68)	-8.4×10^-13	1.2×10^{-3}(1.62)	-8.2×10^-13	6.6×10^{-4}(1.64)	7.1×10^{-13}	2.8×10^{-4}(1.53)
		Kuiper2sample.c	4×10^{-12}	2.8×10^{-4}(1)	6×10^{-12}	7.5×10^{-4}(1)	-8.2×10^-13	4×10^{-4}(1)	5.7×10^{-12}	1.8×10^{-4}(1)
		kuiper.2samp	30.1	6.7×10^{-5}(0.237)	40.2	6.8×10^{-5}(0.0905)	36.6	6.7×10^{-5}(0.169)	18.1	7.1×10^{-5}(0.39)
	500	Kuiper2sample	0	7.5×10^{-3}(1.5)	-6.7×10^-13	0.02(1.74)	-2.1×10^-12	0.014(1.75)	-1.4×10^-12	6×10^{-3}(2)
		Kuiper2sample.c	6.5×10^{-12}	5×10^{-3}(1)	-9.4×10^-12	0.0115(1)	-2.9×10^-11	8×10^{-3}(1)	-4.9×10^-11	3×10^{-3}(1)
		kuiper.2samp	26.3	6.6×10^{-5}(0.0133)	17.7	6.7×10^{-5}(5.8×10^-3)	22.1	6.7×10^{-5}(8.4×10^-3)	20.3	6.8×10^{-5}(0.0227)
	2500	Kuiper2sample	-2.4×10^-12	0.191(2.16)	1.2×10^{-11}	0.482(2.13)	1.2×10^{-12}	0.264(2.13)	-2.9×10^-12	0.11(2.09)
		Kuiper2sample.c	-2.3×10^-10	0.0885(1)	-1.7×10^-10	0.227(1)	1×10^{-11}	0.124(1)	-1×10^-10	0.0525(1)
		kuiper.2samp	6.65	6.7×10^{-5}(7.6×10^-4)	7.85	6.6×10^{-5}(2.9×10^-4)	7.27	6.7×10^{-5}(5.4×10^-4)	4.36	6.8×10^{-5}(1.3×10^-3)
	10000	Kuiper2sample	4.1×10^{-12}	2.42(2.03)	4.7×10^{-12}	6.3(2.13)	5.9×10^{-12}	3.6(2.16)	-2.9×10^-12	1.48(2.14)
		Kuiper2sample.c	-1.1×10^-9	1.19(1)	4.3×10^-9	2.96(1)	-1.3×10^-9	1.67(1)	2.8×10^-10	0.69(1)
		kuiper.2samp	3.35	6.7×10^{-5}(5.6×10^-5)	4.22	7×10^{-5}(2.4×10^-5)	4.17	6.7×10^{-5}(4×10^-5)	2.21	6.7×10^{-5}(9.8×10^-5)
	50000	Kuiper2sample	-1.4×10^-11	53.8(2.15)	-4×10^-12	139(2.15)	0	79.2(2.18)	-2.3×10^-12	32.6(2.18)
		Kuiper2sample.c	-6.5×10^-9	25(1)	-1.7×10^-9	64.7(1)	-1.8×10^-9	36.3(1)	1.6×10^-8	15(1)
		kuiper.2samp	1.69	6.8×10^{-5}(2.7×10^-6)	2.03	6.6×10^{-5}(1×10^-6)	1.72	6.7×10^{-5}(1.9×10^-6)	2.21	6.7×10^{-5}(4.4×10^-6)
	100000	Kuiper2sample	1.8×10^{-11}	208(2.16)	-1.8×10^-11	540(2.18)	-9.1×10^-12	307(2.2)	5.8×10^-13	129(2.22)
		Kuiper2sample.c	-1.2×10^-9	96(1)	-2.5×10^-8	248(1)	2×10^{-8}	139(1)	-2.8×10^-8	58.1(1)
		kuiper.2samp	1.29	6.7×10^{-5}(7×10^-7)	1.23	6.8×10^{-5}(2.7×10^-7)	1.03	6.7×10^{-5}(4.8×10^-7)	1.66	7×10^{-5}(1.2×10^-6)
1	100	Kuiper2sample	-1.3×10^-13	2.9×10^{-4}(1.6)	0	7×10^{-4}(1.57)	1.3×10^{-13}	3.9×10^{-4}(1.57)	-1.3×10^-13	1.7×10^{-4}(1.33)
		Kuiper2sample.c	1.3×10^{-13}	1.8×10^{-4}(1)	-1.4×10^-13	4.4×10^{-4}(1)	-1.4×10^-13	2.5×10^{-4}(1)	0	1.3×10^{-4}(1)
		kuiper.2samp	8.32	8.3×10^{-5}(0.452)	11.6	8.3×10^{-5}(0.186)	13.4	8.4×10^{-5}(0.339)	5.45	8.5×10^{-5}(0.677)
	500	Kuiper2sample	-1.3×10^-13	5×10^{-3}(1.67)	1.3×10^{-13}	0.012(1.71)	-2.6×10^-13	6.5×10^{-3}(2.17)	-2.7×10^-13	3.5×10^{-3}(2.33)
		Kuiper2sample.c	-3.9×10^-13	3×10^{-3}(1)	0	7×10^{-3}(1)	2.7×10^{-13}	3×10^{-3}(1)	-5.3×10^-13	1.5×10^{-3}(1)
		kuiper.2samp	5.61	8.4×10^{-5}(0.028)	6.05	8.4×10^{-5}(0.012)	5.05	8.4×10^{-5}(0.028)	8.17	8.2×10^{-5}(0.055)
	2500	Kuiper2sample	0	0.1(2.5)	-2.6×10^-13	0.251(2.61)	-8.8×10^-13	0.139(2.5)	0	0.057(2.43)
		Kuiper2sample.c	-4.9×10^-12	0.04(1)	-1×10^-11	0.096(1)	-5×10^-13	0.0555(1)	-5.8×10^-12	0.0235(1)
		kuiper.2samp	2.13	8.4×10^{-5}(2.1×10^-3)	2.96	8.4×10^{-5}(8.8×10^-4)	2.69	8.4×10^{-5}(1.5×10^-3)	1.42	8.3×10^{-5}(3.6×10^-3)
	10000	Kuiper2sample	-8.5×10^-13	1.21(2.09)	2.4×10^-13	3.08(2.37)	-1.1×10^-12	1.75(2.43)	-1.8×10^-12	0.72(2.25)
		Kuiper2sample.c	-2.8×10^-11	0.58(1)	-1.9×10^-11	1.3(1)	-3×10^-11	0.72(1)	-6×10^-12	0.32(1)
		kuiper.2samp	1.09	8.8×10^{-5}(1.5×10^-4)	1.43	8.2×10^{-5}(6.3×10^-5)	1.04	8.4×10^{-5}(1.2×10^-4)	0.726	8.4×10^{-5}(2.6×10^-4)
	50000	Kuiper2sample	-1.7×10^-12	25.6(2.23)	6.9×10^-12	65.4(2.15)	2.6×10^{-12}	37.1(2.19)	1.7×10^{-12}	15.4(2.17)
		Kuiper2sample.c	4×10^{-11}	11.5(1)	-1.5×10^-10	30.3(1)	-2.4×10^-10	16.9(1)	-6.9×10^-12	7.1(1)
		kuiper.2samp	0.762	8.3×10^{-5}(7.3×10^-6)	0.573	8.3×10^{-5}(2.8×10^-6)	0.529	8.4×10^{-5}(5×10^-6)	0.595	8.5×10^{-5}(1.2×10^-5)
	100000	Kuiper2sample	5.8×10^{-12}	97.2(2.17)	-1.8×10^-12	251(2.18)	-1.1×10^-12	142(2.2)	4.9×10^-13	59(2.18)
		Kuiper2sample.c	1.1×10^{-10}	44.8(1)	-1.1×10^-10	115(1)	-7.5×10^-10	64.7(1)	2.9×10^-10	27(1)
		kuiper.2samp	0.48	8.7×10^{-5}(1.9×10^-6)	0.463	8.3×10^{-5}(7.2×10^-7)	0.398	8.4×10^{-5}(1.3×10^-6)	0.603	8.3×10^{-5}(3.1×10^-6)

Chapter 3

On a One Sample Goodness-of-Fit Test Based on the Hausdorff Metric

This chapter is based on the paper:

Dimitrina S. Dimitrova, Yun Jia and Vladimir K. Kaishev (2025). On a One Sample Goodness-of-Fit Test Based on the Hausdorff Metric. *submitted*.

Abstract

We consider the Hausdorff metric and its use in measuring the distance between an empirical and a theoretical cumulative distribution function (cdf). We propose a corresponding one-sample Hausdorff goodness-of-fit test statistic, the H test, give its geometric interpretation, and a method to evaluate it. We show that its exact and asymptotic distributions can be expressed correspondingly as rectangle probability and as double boundary crossing probability with respect to a Brownian bridge. Efficient numerical methods for computing the distributions, the exact Bahadur slope, and asymptotic power of the H test are also provided.

We also show that the p-values of the H test are not invariant under scale transformation. Investigating theoretically this scale dependence, we find that by appropriately selecting the scale coefficient, the power of the H test can be controlled and optimized. This is an important feature which other tests such as Kolmogorov-Smirnov (KS), Cramer von Mises (CvM) and the Anderson-Darling (AD) do not possess. In particular, based on synthetic and real data examples, we demonstrate that when testing goodness-of-fit in the tail, the power and tail sensitivity of the scale-tuned H test is higher than the power of the KS, CvM and AD tests. All these properties make the H test a competitive alternative to existing goodness-of-fit tests.

3.1 Introduction

Consider the classical, one-sample goodness-of-fit problem of whether a random sample comes from a pre-specified hypothetical (null) distribution. There are many goodness-of-fit test statistics in the literature among which the classical Kolmogorov-Smirnov (KS), the Kuiper, the Cramer von Mises (CvM), the Anderson-Darling (AD) and the Wasserstein (W) test more recently considered in del Barrio et al. (1999) and del Barrio et al. (2000). The latter tests have gained great popularity and have been widely applied in almost any field where data is collected and analysed such as, astronomy (McQuillan et al., 2013), social sciences (Salman et al., 2015), pattern recognition (Alzubaidi and Kalita, 2016), machine learning (Gretton et al., 2012) etc., to name only a few.

These and other existing tests have different properties and there is no "best" test that suits all purposes and posesses all the best properties. For example, the very popular KS test, based on the supremum distance, is readily understood graphically, is easy to evaluate and is distribution-free when the null is continuous. Furthermore, recently Dimitrova et al. (2020) provided efficient means of computing the KS p-values assuming arbitrary continuous, discrete, or mixed null distribution, which makes the KS test applicable beyond just the continuous case. At the same time, the KS statistic is less sensitive in the tails, and has in general lower power (see e.g. Mason and Schuenemeyer, 1983; Feigelson and Babu, 2020). This makes the KS test less efficient, especially for comparing tails, which is very important in extreme value applications and related inference.

The AD and CvM tests, based on the L^2-distance, are also distribution-free and have high power but have not been introduced for discrete or mixed null distributions. The W test, based on the L^2-Wasserstein distance, also has high power but is difficult to evaluate numerically and is not distribution free, which hinders its practical use.

As noted, each test is defined using a particular metric to measure the distance between the underlying cumulative distribution functions. The definition of the distance metric determines the properties of the test, in particular the evaluation of the test, its distributions and power. All these considerations lead to the conclusion that there is still scope for applying alternative distance metrics, leading to the

construction of new test statistics and the need to investigate their related properties.

The aim of this chapter is to explore how the Hausdorff metric, introduced by Hausdorff (1914) to measure the distance between sets, can be applied to measure the distance between (empirical) cumulative distribution functions on the real line, with the purpose of introducing a corresponding Hausdorff goodness-of-fit test statistic. The Hausdorff distance has been considered by Sendov and Beer (2012) within the context of approximation theory, and in machine learning by Huttenlocher et al. (1993), Li et al. (2017), Karimi and Salcudean (2020), Zhao et al. (2021), to name only a few of the papers in this stream of literature.

To the best of our knowledge, the Hausdorff metric has not been previously considered to measure the distance between cdfs, with the only exceptions of Rachev (1984) and Bloch and Atif (2016). While Rachev (1984) shows that the Levy and Hausdorff distances between two cdfs are expressed in a similar way (cf. Theorem 1 therein), Bloch and Atif (2016) give an equivalent expression by exploiting the link between the Hausdorff distance and morphological dilation (cf. expression (11) therein). However, the latter expression is in general terms, leaving open the question of how to efficiently compute the Hausdorff distance between two cumulative distribution functions.

Here, we provide an answer to this question. Secondly, we use the Hausdorff distance to construct a corresponding Hausdorff goodness-of-fit test statistic, develop an efficient method to calculate its distributions and investigate some of its most important properties. It will be convenient to refer to the latter statistic as the H (test) statistic.

Let us note that while the H test is invariant to location shifting, it is not invariant under scale transformation, in contrast to other tests which are both location and scale invariant (cf. KS, AD, etc.). We show that the lack of scale invariance has some desirable effects on the H test. More precisely, we demonstrate that by appropriately selecting the scale parameter it is possible to tune the H statistic in such a way that its power is maximized. This makes the newly proposed Hausdorff test a powerful competitor to other tests when comparing tails of distributions, in the context of probability theory and statistical inference for extreme values.

Our major contributions can be summarized as follows. First, in Lemma 3.6

we show that the general definition of the Hausdorff distance simplifies significantly when applied to cumulative distribution functions. Based on this and on Lemmas 3.8 and 3.9, we provide an efficient method to compute it. The pseudocode of the method is described in Algorithm 1. We also establish some useful inequalities between the Hausdorff and Kolmogorov-Smirnov distances and their corresponding distributions (see Theorem 3.11 and Corollary 3.12). The known equivalence between the Hausdorff and Lévy metrics is recalled and used to establish some useful results with respect to the Hausdorff distance.

Second, we propose to use the Hausdorff distance as a goodness-of-fit measure and, by applying a suitable coordinate transformation, we show in Theorem 3.17 that the H statistic can be expressed alternatively as the supremum of an absolute difference, resembling the form of the Kolmogorov-Smirnov statistic. Using this form, in Theorem 3.19 we are able to express, the exact distribution of the H statistic as a Steck (1971) rectangle probability which can be efficiently computed (with complexity $O(n^2 \log n)$) using the exact KS-FFT based method proposed by Dimitrova, Kaishev and Tan (2020). In Theorem 3.21 we give the asymptotic distribution expressed as a double boundary non-crossing probability with respect to a Brownian bridge. Closed form expressions for the asymptotic distributions in the special cases of exponential and uniform nulls are obtained in Corollary 3.22 and Theorem 3.23. The latter establishes a direct connection between the H test and the KS test and their distributions.

Third, we take advantage of the scale dependence of the H statistic. By specifying its confidence band, we show in Proposition 3.27 that its power admits a representation as a double boundary non-crossing probability, where the boundaries and the band between them depend on the scale parameter. Therefore, based on Theorem 3.29 and Corollaries 3.30 and 3.31, we show that under the assumption of concavity/convexity of the null or its right/left tails, the power can be locally controlled, making H body or right/left tail sensitive by appropriately selecting the scale parameter. In particular Theorem 3.29 suggests that if the null has a concave right tail, and if it deviates from the true therein, the H test would become more right-tail sensitive as the scale parameter decreases. Since most commonly used null distributions such as Exponential, Normal, Pareto, Lognormal, Gamma, etc, have

concave right tails, in this case an optimal power of the $\mathcal{H}_n(\sigma)$ test is achieved by appropriately selecting a value of the scale parameter. In (3.3.11) we propose such an optimal choice which solely depends on the null distribution and is motivated by the geometric interpretation of the transformed H statistic, provided by Proposition 3.34. Proposition 3.35 shows that the choice (3.3.11) is invariant under further rescaling thus ensuring that the issue of scale dependence of the proposed H test is eliminated.

Last but not least, we provide some insights on the asymptotic power of the proposed H test. In Theorem 3.41, we give a computable expression for the exact Bahadur slope of H and use it to evaluate the Bahadur efficiency of H, relative to the efficiency of the KS test, based on an example where the alternative is a spliced distribution. The results, summarized in Table 3.2, show that asymptotically the Hausdorff test is more efficient than the KS test when the null and alternative differ in the tail. In Theorem 3.39, we provide an expression of the asymptotic power of the H test under an appropriate contiguous alternative.

To emphasize the practical importance of the proposed H test, we provide simulated examples and a real data example which demonstrate that for finite sample sizes the H test has higher power than the KS, CvM and AD tests when samples deviate from the null distribution in the tail. The real data example, involving Lloyd's aviation insurance losses, also illustrates the higher tail sensitivity of the H test compared to the classical KS, CvM and AD tests. As can be seen from Table 3.4, H is the only test that rejects the null hypothesis, capturing the tail discrepancy between the null distribution and the sample, indicated by the corresponding QQ plot in Fig 3.12.

This chapter is organized as follows. In Section 3.2, we recall the general definition of the Hausdorff metric and introduce the H goodness-of-fit statistic. We further provide methods to compute the H statistic and its exact and asymptotic distributions. Connections between the H and KS tests and their distributions are also established. In Section 3.3, we show that the sensitivity of the H test can be controlled locally by varying the scale parameter and propose a rule to select it so that the power is optimized in the area (e.g. body or left/right tail) where it is of interest to test the deviation between the null and the sample. In Section 3.4, we

give some results on the asymptotic power of the Hausdorff and KS statistics applying the Bahadur exact slope and efficiency. In Section 3.5, we provide simulated and real-data examples to demonstrate the advantage of the proposed H test in extreme value goodness-of-fit testing. In Section 3.6, we provide a summary and discussion of our findings.

Further numerical examples are given in Appendix 3.B. The proofs of all the results that appear in the chapter are given in Appendix 3.A.

3.2 The Hausdorff Goodness-of-fit Test Statistic

Given a random sample $\boldsymbol{X}_n = \{X_1, \ldots, X_n\}$, where X_i, $i = 1, \ldots, n$, are i.i.d. copies of the random variable, X, defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, with (unknown) cumulative distribution function $F_0(x)$. We want to test the null hypothesis, H_0, that the sample $\boldsymbol{X}_n$ comes from a pre-specified (null) distribution, with distribution function $F(x)$, i.e., $H_0 : F_0(x) = F(x)$, for all $x \in \mathbb{R}$, against the alternative, $\rho_\infty : F_0(x) \neq F(x)$, for at least one x. For convenience, we assume $F(x)$ is continuous, however, most of our results are either directly applicable for the case of F with jumps, or applicable with slight adjustments. Denote by $F_n(x)$, $x \in \mathbb{R}$, the empirical cumulative distribution function, (ecdf), corresponding to the sample $\boldsymbol{X}_n$, i.e.

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}(X_i \leq x), \tag{3.2.1}$$

where $\mathbb{1}(\cdot)$ is the indicator function. In the sequel, it will be convenient to interchangeably use the notation $F_n(x) \equiv F_n(x, \omega) \equiv F_n$, where $\omega \in \Omega$ explicitly indicates that the empirical cdf depends on the random sample $\boldsymbol{X}_n(\omega)$.

3.2.1 Background on the Hausdorff Metric

In order to introduce our test statistic based on the Hausdorff metric, we need to recall the following definition of a distance measure between two points, A and B, on the plane, $\mathbb{R}^2$.

Definition 3.1. The function $\rho(A, B)$ is a distance measure between two points $A, B \in \mathbb{R}^2$ iff it satisfies the following conditions:

1. $\rho(A, B) \geq 0$, for every pair of points A and B.

2. $\rho(A, B) = 0$, iff the two points A and B coincide.

3. Symmetry: $\rho(A,B) = \rho(B,A)$.

4. Triangle inequality: $\rho(A,B) + \rho(B,C) \geq \rho(A,C)$.

Example 3.2. Examples of such a distance measure are given by the following functions:

$$\rho_\infty(A,B) = \max\{|x_A - x_B|, |y_A - y_B|\} \tag{3.2.2}$$

$$\rho_1(A,B) = |x_A - x_B| + |y_A - y_B| \tag{3.2.3}$$

$$\rho_2(A,B) = \sqrt{(x_A - x_B)^2 + (y_A - y_B)^2}, \tag{3.2.4}$$

where x_A, x_B and y_A, y_B are correspondingly, the x and y coordinates of the points A and B.

It can be directly verified that, all functions, h_i, $i = 1,2,3$, in Example 3.2, satisfy conditions 1-4 of Definition 3.1. Their geometric interpretations are given in Figure 3.1.

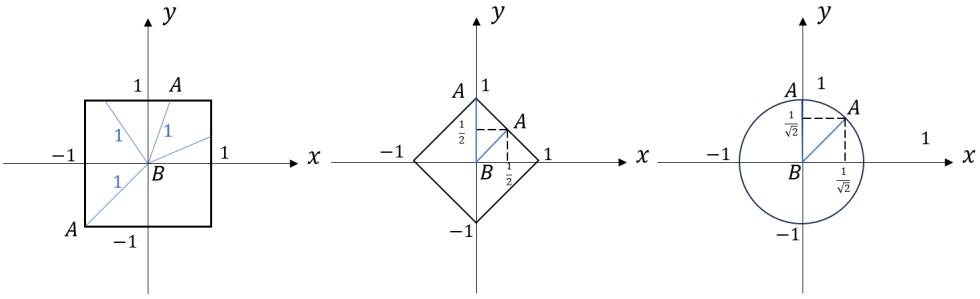


Figure 3.1: All points, A, located at unit distance from $B(0,0)$, with respect to $\rho_i(A,B)$, $i = \infty, 1, 2$, defined in Example 3.2, form correspondingly: a) a square; b) a rombus and c) a circumference.

In what follows, it will be conveneint to work with definition, ρ_∞, which is illustrated in panel a) of Figure 3.1, where the point A can be any point on the square with sides 2, and center, the point, B. In other words, the set of all points, $\{A : \rho_\infty(A,B) = 1\}$, coincides with the contour of the square, which therefore can be viewed as the circumference with respect to ρ_∞, with radius 1 and center, the point B.

We are now in a position to recall the definition of the Hausdorff distance between sets. For the purpose, assume, $\rho(A,B)$ is an arbitrary measure of distance between two points A and B, on $\mathbb{R}^2$. Let also $\mathcal{A}$ and $\mathcal{B}$ be two arbitrary sets on the

plane, $\mathbb{R}^2$, bounded with respect to ρ. We can now define the Hausdorff distance as follows.

Definition 3.3. The Hausdorff distance, $H_\rho(\mathcal{A},\mathcal{B})$ between the sets $\mathcal{A}$ and $\mathcal{B}$ is defined as

$$H_\rho(\mathcal{A},\mathcal{B}) = \max\left\{\sup_{A\in\mathcal{A}}\inf_{B\in\mathcal{B}}\rho(A,B), \sup_{B\in\mathcal{B}}\inf_{A\in\mathcal{A}}\rho(A,B)\right\}. \tag{3.2.5}$$

For a proof that $H_\rho(\mathcal{A},\mathcal{B})$ is a distance metric, we refer to Sendov and Beer (2012).

3.2.2 Properties of the Hausdorff Metric Applied to Cumulative Distribution Functions

In order to apply the Hausdorff metric $H_\rho(\mathcal{A},\mathcal{B})$ to measure the distance between the distribution functions, $F(x)$ and $F_n(x)$, $x\in\mathbb{R}$, we need to appropriately define the sets $\mathcal{A}$ and $\mathcal{B}$ from Definition 3.3. We replace them with the planar curve (i.e., set) analogues, F^c and F_n^c, of the (right-continuous) cdfs, $F(x)$ and $F_n(x)$, completed by vertical segments, corresponding to the jumps at their points of discontinuity. Thus, for the planar curve $F_n^c(\omega)$, which is a closed and connected subset in $\mathbb{R}^2$, we have

$$F_n^c(\omega) = \{(x,y) : F_n(x-,\omega) \le y \le F_n(x,\omega), \forall\, x\in\mathbb{R}\}. \tag{3.2.6}$$

A more general definition of the planar curve is given in (4.2.6) of Chapter 4.

Note that, if $F(x)$ is continuous, its corresponding planar curve F^c, defined as in (3.2.6), coincides with its graph, $gr(F) = \{(x,F(x)) : x\in\mathbb{R}\}$, which is a closed and connected set in $\mathbb{R}^2$. Therefore, for simplicity and symmetry, we use the notation F^c for the planar curve of the null distribution F.

Remark 3.4. When the underlying distribution is discrete or mixed, i.e., when $F(x)$ has jumps, its corresponding planar curve, F^c, is defined as in (3.2.6), so as to fill in the missing parts at the jumps.

In order to test the null hypothesis H_0, introduced in Section 3.2, we propose to use the Hausdorff distance, $H_\rho(F^c,F_n^c)$, between the planar curves F^c and F_n^c, that correspond to the cdfs, F and F_n. Let us note that, $H_\rho(F^c,F_n^c)$, is a well defined goodness-of-fit test statistic, for an arbitrary choice of the distance, ρ, in particular, for $\rho=\rho_i$, $i=\infty,1,2$, as in Example 3.2. To the best of our knowledge, using

$H_\rho(F^c, F_n^c)$, to test the null hypothesis, H_0, has not been previously considered, under any choice of $\rho = \rho_i$, $i = \infty, 1, 2$. The Hausdorff distance under ρ_2 has been considered by Popov (1999) but with respect to arbitrary functions.

As a measure of the distance from a point, on the curve, F^c to any point on F_n^c, we have chosen the metric ρ_∞ defined in (3.2.2), i.e. we take $\rho = \rho_\infty$. This is motivated by the following reasons. Firstly, since, F_n^c is a staircase curve, and since the metric H_{ρ_∞} is the side of the largest square inscribed between the curves F^c and F_n^c, (see Lemma 3.8), inscribing squares is more natural and computationally appealing than inscribing other shapes, e.g. rombuses or circles as in Fig 3.1. Secondly, the choice, $\rho = \rho_\infty$, ensures that the Hausdorff metric H_{ρ_∞} coincides with the Lévy metric, cf (3.2.14). The latter is yet another metric used to measure the distance between two probability distribution functions. It has some useful properties, which we consider later in this section. Lastly, if $\rho = \rho_\infty$, the Hausdorff distance between the graphs, $gr(F)$ and $gr(F_n)$, of the cdfs F and F_n, and between their corresponding planar curves coincide, as we show in the following lemma. These geometric conveniences lead to an efficient method of computing it (c.f Algorithm 1) and elegant theoretical results, given in Section 3.2. For brevity, we will drop the subscript ρ_∞ from $H_{\rho_\infty}(F^c, F_n^c)$ and write $H(F^c, F_n^c)$.

Lemma 3.5. *We have,* $H(gr(F), gr(F_n)) = H(F^c, F_n^c)$.

Since there is a one-to-one correspondence between the functions F and F_n and their graphs, from Lemma 3.5 it follows that, $H(F, F_n)$ and $H(F^c, F_n^c)$ are equivalent, so one can view H as a functional, i.e.,

$$H(F, F_n) = \max\left\{ \sup_y \inf_x \max\left(|x-y|, |F(x) - F_n(y)|\right), \right. \\ \left. \sup_x \inf_y \max\left(|x-y|, |F(x) - F_n(y)|\right) \right\}. \tag{3.2.7}$$

We will refer to the Hausdorff distance between F_n and F as the proposed H test statistic denoted by $H(F^c, F_n^c)$.

The following lemma is important since it shows that the general expression (3.2.5) for the Hausdorff distance simplifies significantly when applied to measuring the distance between cdfs.

Lemma 3.6. *The Hausdorff distance*

$$H(F^c, F_n^c) = \sup_{A \in F_n^c} \inf_{B \in F^c} \rho_\infty(A, B) = \sup_{B \in F^c} \inf_{A \in F_n^c} \rho_\infty(A, B).$$

where $\rho_\infty = \max\{|x_A - x_B|, |y_A - y_B|\}$.

As we argued in the introduction, (and demonstrate in the sequel), the test statistic $H(F^c, F_n^c)$ has some nice properties that make it a competitive alternative to exisitng tests, such as Kolmogorov-Smirnov, Anderson-Darling, Wasserstein, etc. The H test is more flexible and in general, captures differences between the cdfs F_n and F, better than purely ordinal tests, such as KS, CvM and AD tests. This is because, $H(F^c, F_n^c)$ depends on both the x and y coordinates of the curves F_n^c and F^c, as can be seen from the definition of the distance metric ρ_∞, given in Example 3.2-(2). This is in contrast to, e.g. the Kolmogorov-Smirnov statistic which depends only on the y-axis, by definition.

Remark 3.7. In general, for other specifications of ρ, e.g., for $\rho = \rho_1, \rho_2$ defined in (3) and (4) of Example 3.2, the statistic $H_\rho(F^c, F_n^c)$, also depends on both the x and y coordinates of the curves F_n^c and F^c, therefore is also better suited to capture the differences between them, than the purely ordinal tests.

In order to express the Hausdorff test statistic $H(F^c, F_n^c)$, and formulate some important results, we need to introduce the following notation.

Let, $\{x_1, \ldots, x_n\}$ denote a realization of the random sample $\boldsymbol{X}_n = \{X_1, \ldots, X_n\}$, such that $x_1 < x_2 < \cdots < x_n$. The empirical cdf $F_n(x)$ and its planar counterpart, F_n^c with vertexes, $A_1, A_2, \ldots, A_{2n}$, having coordinates

$$A_{2i} = \Big(x_i, F_n(x_i)\Big);\ A_{2i-1} = \Big(x_i, F_n(x_{i-1})\Big)\ i = 1, 2, \ldots, n, \tag{3.2.8}$$

where $F_n(x_0) \equiv 0$, for some $x_0 < x_1$. Both, $F_n(x)$ and F_n^c are illustrated in Figure 3.2, for the case, $n = 3$.

Let F^c divide the plane into two open sets, denoted by $\mathcal{U}_{F^c}$ and $\mathcal{L}_{F^c}$, and correspondingly referred to as the (strict) epigraph and hypograph of F^c. Simliarly, F_n^c divides $\mathbb{R}^2$ plane into $\mathcal{U}_{F_n^c}$ and $\mathcal{L}_{F_n^c}$. We are now in a position to formulate some important lemmas, starting with Lemma 3.8, which establishes a relation between

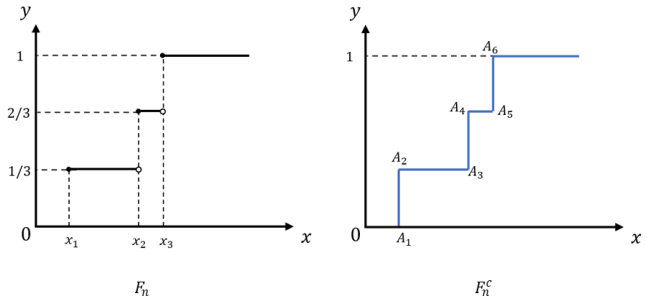


Figure 3.2: F_n is function whereas F_n^c is planar curve

the Hausdorff distance and a square that can be fitted between the curves, F_n^c and F^c. In particular, the side of the maximal such square is the Hausdorff distance $H(F^c, F_n^c)$.

Lemma 3.8. *If a square* $S(P,d) = \{P_1 \in \mathbb{R}^2 : \rho_\infty(P_1, P) \leq d/2\}$, *with side* d *and center at the point* P, *can be inserted between the curves* F^c *and* F_n^c, *so that* $S(P,d)$ *does not overlap with the sets* $\mathcal{U}_{F^c} \cap \mathcal{U}_{F_n^c}$ *and* $\mathcal{L}_{F^c} \cap \mathcal{L}_{F_n^c}$, *then for the Hausdorff distance, we have*

$$H(F^c, F_n^c) \geq d. \tag{3.2.9}$$

Furthermore,

$$H(F^c, F_n^c) = \sup\{d\}, \tag{3.2.10}$$

where the supremum is taken over all

$$S(P,d) : S(P,d) \cap \{(\mathcal{U}_{F_n^c} \cap \mathcal{U}_{F^c}) \cup (\mathcal{L}_{F_n^c} \cap \mathcal{L}_{F^c})\} = \emptyset, \tag{3.2.11}$$

or equivalently, over all

$$S(P,d) : S(P,d) \subset \mathcal{G}, \tag{3.2.12}$$

where $\mathcal{G} = \mathbb{R}^2 / \{(\mathcal{U}_{F^c} \cap \mathcal{U}_{F_n^c}) \cup (\mathcal{L}_{F^c} \cap \mathcal{L}_{F_n^c})\}$, *is the area between the two planar curves* F^c *and* F_n^c.

Lemma 3.8 is important as it provides a geometric interpretation of the distance, $H(F^c, F_n^c)$, and gives insight on how to numerically compute $H(F^c, F_n^c)$, and link it to the Lévy metric. The latter is used to measure the distance between two probability distribution functions. For any two arbitrary cdfs $F(x)$ and $G(x)$, the Lévy metric

is defined as (see e.g. Zolotarev, 2011)

$$\rho_{Levy}(F,G) = \inf\{\epsilon : F(x-\epsilon)-\epsilon \leq G(x) \leq F(x+\epsilon)+\epsilon\}. \tag{3.2.13}$$

It is important to note that both H and the Lévy metric, ρ_{Levy} have a coincident geometric interpretation as the side of the largest square inscribed between the corresponding planar curves, as shown in Lemma 3.8, for H and as illustrated by Zolotarev (2011), for ρ_{Levy}. Therefore, the Hausdorff metric H and the Lévy metric, coincide, i.e.

$$H(F^c,G^c) = \rho_{Levy}(F,G), \tag{3.2.14}$$

where G^c is the planar curve of $G(x)$. This fact has also been highlighted in Theorem 1^* of Rachev (1984). Therefore, the properties of the Lévy metric are also valid for the metric H, as summarized by the Lemma 3.46, which is a restatement of properties (2) (6) and (7), summarized in Zolotarev (2011) with respect to the Lévy metric. For further upper bounds with respect to the Lévy metric, we refer to Zolotarev (1971). We should highlight that Alexander (1974) has proposed the goodness-of-fit statistic based on the Lévy metric, however, the Alexander (1974)'s statistic shares the same property as KS. For more discussion, we refer to Remark 3.24.

We can now address the question of how to compute $H(F^c,F_n^c)$, by specifying result (3.2.10), of Lemma 3.8. Denote by, $\mathcal{B}_{loc} = [(\{A_{2i-1}\}_{i=1}^n \cap \mathcal{L}_{F^c}) \cup (\{A_{2i}\}_{i=1}^n \cap \mathcal{U}_{F^c})\}] = \{B_1, B_2, \ldots, B_\nu\}$, the part of the vertices of the curve F_n^c that are locally farthest from F^c. The following proposition gives an explicit expression for the Hausdorff distance, based on which we develop numerical methods to compute the value of the statistic $H(F^c,F_n^c)$.

Lemma 3.9. *Let, $\mathcal{L}_l$, $l = 1,2,\ldots,\nu$ be parallel straight lines, correspondingly passing through each of the vertices, B_l, $l = 1,2,\ldots,\nu$, in such a way that they cross the x-axis, at an angle of $3\pi/4$. Denote by E_l, $l = 1,2,\ldots,\nu$ the points of intersection of the lines, $\mathcal{L}_l$ with the planar curve F^c and consider the distances $\rho_\infty(B_l,E_l)$, $l = 1,2,\ldots,\nu$. Then we have*

$$H(F^c,F_n^c) = \max\{\rho_\infty(B_l,E_l), l = 1,\ldots,\nu\}. \tag{3.2.15}$$

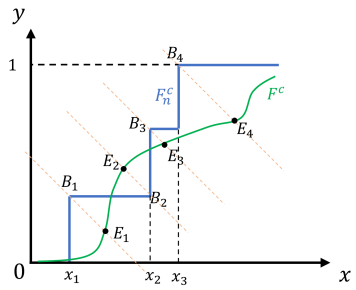


Figure 3.3: Illustration of the lines $\mathcal{L}_l$, and the points B_l and E_l, $l=1,2,3,4$, from Lemma 3.9.

Lemma 3.9 is geometrically illustrated in Fig. 3.3. Equation (3.2.15) is essential for computing the value of $H(F^c,F_n^c)$. In particular, when F is continuous, the computation is straightforward, as illustrated in Algorithm 1. Denote by x_{B_l} and y_{B_l}, the x and y coordinates of the points B_l, $l=1,\ldots,\nu$, respectively. Then it is easy to find all the crossing points, $E_l=(x_l^*,F(x_l^*))$ of the straight line $\mathcal{L}_l=\{(x,y): x+y=x_{B_l}+y_{B_l}\}$ with the curve F^c, where x_l^*, $l=1,\ldots,\nu$ are the unique solutions of the equations in line 5. The latter uniqueness is guaranteed by the continuity of $F(x)$. The value of $H(F^c,F_n^c)$ is then found via the maximization in line 6.

Algorithm 1: Computing the value of $H(F^c,F_n^c)$

Data: $F,x_1,\ldots,x_n$

Result: $y=H(F^c,F_n^c)$

1 $\mathcal{B}_{loc}\leftarrow\{(x_i,\frac{i-1}{n}):F(x_i)\geq\frac{i-1}{n}\}\cup\{(x_i,\frac{i}{n}):F(x_i)<\frac{i}{n}\}$;

2 $\nu\leftarrow|\mathcal{B}_{loc}|$;

3 $y\leftarrow 0$;

4 **for** $l=1;\ l\leq\nu;\ l=l+1$ **do**

5 $\quad x_l^*\leftarrow$ the solution of $x+F(x)=x_{B_l}+y_{B_l}$;

6 $\quad y\leftarrow\max(y,|x_l^*-x_{B_l}|)$;

7 **end**

Remark 3.10. Lemma 3.9 is also true for distribution functions $F(x)$ with jumps. However, in the latter case, equation $x+F(x)=x_{B_l}+y_{B_l}$ may not have a solution. since the line $\mathcal{L}_l$ may cross $F(x)$ at some of the jump discontinuities. Therefore, to find the points E_l, $l=1,\ldots,\nu$, one may need to accordingly adjust line 5 in Algorithm 1.

Next, based on Lemma 3.9 we state Theorem 3.11, which is central since it gives

a connection between $H(F^c, F_n^c)$ and the one sample Kolmogorov-Smirnov statistic, $\mathcal{D}_n$ defined as

$$\mathcal{D}_n = \rho_D(F, F_n) := \sup_{-\infty<x<+\infty} |F_n(x) - F(x)|. \tag{3.2.16}$$

Theorem 3.11. *For any distribution $F(x)$, we have*

$$H(F^c, F_n^c(\omega)) \leq \rho_D(F, F_n(\omega)) \tag{3.2.17}$$

for every $\omega \in \Omega$.

Let us note that the following corollary with respect to the cdfs of H and KS is a direct consequence of Theorem 3.11.

Corollary 3.12. *For any $q \in [0,1]$, $\mathbb{P}(H(F^c, F_n^c) > q) \leq \mathbb{P}(\mathcal{D}_n > q)$.*

3.2.3 The Exact and Asymptotic Distributions of the One-sample Hausdorff Statistic H

In this section, we provide means for efficient computation of exact and asymptotic distributions of the statistic H. Our approach is based on performing an appropriate transformation of the planar curves F^c and F_n^c into a new coordinate system. We show that, by applying the transformation, the value of the statistic H can be expressed alternatively in the form of the supremum of an absolute difference, analogous to the form of the Kolmogorov-Smirnov statistic. Using this form, we are able to express the exact cdf as a rectangle probability which can be efficiently computed via the exact KS-FFT based method proposed by Dimitrova, Kaishev and Tan (2020). We further extend the latter expression and obtain the asymptotic cdf.

Before introducing the new coordinate system, it would be useful to provide some auxiliary properties of the largest square, $S(P_0, d_0)$, with side, $d_0 = H(F^c, F_n^c)$, which can be fitted in $\mathcal{G}$, i.e. $S(P_0, d_0) \subset \mathcal{G}$, see Figure 3.4. It is easy to see that the upper-left and lower-right vertices of the square $S(P_0, d_0)$, lie on either of the curves F^c and F_n^c. This is formally stated by the following lemma.

Lemma 3.13. *Denote by $d_0 = H(F^c, F_n^c)$, and if $\exists\ P_0 = (x_0, y_0) \in \mathbb{R}^2$ such that $S(P_0, d_0) \subset \mathcal{G}$, then the upper-left and lower-right vertices of $S(P_0, d_0)$ lie on one of*

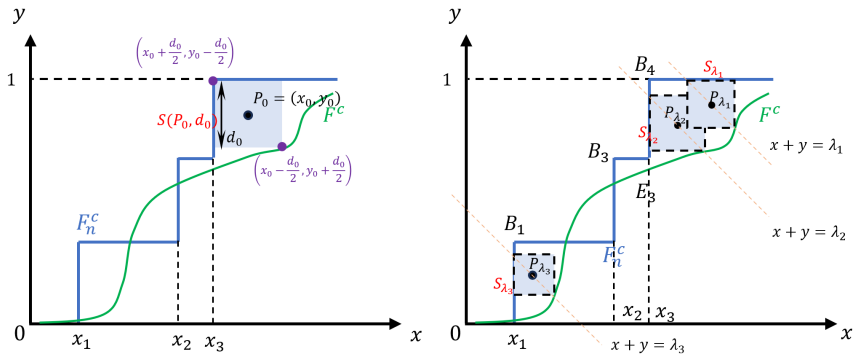


Figure 3.4: Left panel: Illustration of Lemma 3.13 - the upper-left and lower-right vertices of $S(P_0,d_0)$ lie on the curves F^c or F_n^c; Right panel: Illustration of the class $\{S_\lambda\}$

the curves, F^c, F_n^c*, i.e.,*

$$(x_0-d_0/2, y_0+d_0/2), (x_0+d_0/2, y_0-d_0/2) \in F^c \cup F_n^c$$

Hence, to find the largest square, one only needs to search within the class of squares, whose upper-left and lower-right vertices lie on the curves F^c and F_n^c, which we denote as $\mathcal{S}_\lambda$. Specifically, each element $S_\lambda \in \mathcal{S}_\lambda$ indexed by λ, is the square with line $\{(x,y): x+y=\lambda\}$ passing through its center, and satisfying $S_\lambda \subset \mathcal{G}$, $S_\lambda \cap F^c \neq \emptyset$, $S_\lambda \cap F_n^c \neq \emptyset$, as illustrated in the right panel of Figure 3.4. Formally, we have

Lemma 3.14. *Denote by* $P_\lambda = (x_\lambda, y_\lambda)$ *the center of the square* S_λ *and by* d_λ *the length of its side. Then*

$$H(F^c, F_n^c) = \sup_\lambda d_\lambda,$$

where λ *is the index for the line* $\{(x,y): x+y=\lambda\}$ *passing through the center of* S_λ*, i.e.* $\lambda = x_\lambda + y_\lambda$.

The statement of Lemma 3.14 is equivalent to (2.2) in Rachev (1981), formulated with respect to the Lévy metric. For the sake of consistency, in Appendix 3.A we give its proof for the Hausdorff metric.

Remark 3.15. Lemmas 3.13 and 3.14, are also true for arbitrary right continuous nondecreasing functions F' and G', as substitutes of F and F_n, respectively.

Lemma 3.14 is a further generalization of Lemma 3.9, which is needed to express the Hausdorff metric as an appropriate supremum under the transformed coordinate system.

Now, let us introduce an operator to transform the coordinates of points in the plane. Denote by $T:\mathbb{R}^2\to\mathbb{R}^2$ a 2-dimensional coordinate transformation operator such that, for any point $P=(x,y)$, $T(P)=(x+y,y)$. It is easy to verify that the operator T is a bijection and its inverse operator is then $T^{-1}(u,v)=(u-v,v)$. For an arbitrary subset $\mathcal{A}$ in the plane xOy, denote by $T(\mathcal{A})=\{T(P):P\in\mathcal{A}\}$ its corresponding image set on the plane uOv. Before formally introducing the alternative expression of $H(F^c,F_n^c)$ under T, we give the following descriptions for the image sets of some important subsets.

Example 3.16. (i). For any square $S(P,d)$ in the plane xOy, the corresponding image set $T(S(P,d_0))$ is a parallelogram with both its base and height equal to d_0 and one of the diagonals vertical, as shown in Figure 3.5.

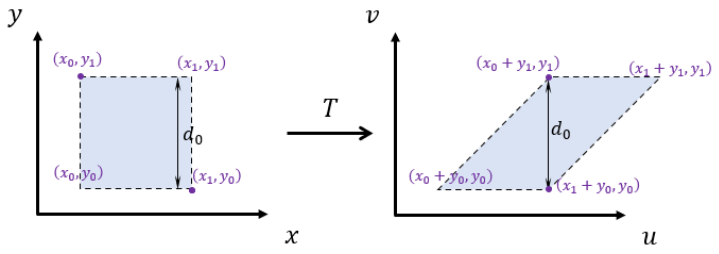


Figure 3.5: Graphical illustration of $S(P,d_0)$ and $T(S(P,d_0))$

(ii). Let F, be a continuous cdf. Let the function K have the following form

$$K(u)=F(\eta^{-1}(u)), \tag{3.2.18}$$

where $\eta^{-1}(u)$ is the inverse function of $\eta(x)=x+F(x)$, $x\in\mathbb{R}$. Then the graph of the parametric equation $v=K(u)$ coincides with the image set of the planar curve F^c, i.e. $T(F^c)=\{(u,v):v=K(u),\text{ for every real } u\}$.[1]

(iii). Similarly as in (ii), denote by $v=K_n(u)$ the parametric equation corresponding to the curve $T(F_n^c)$ on the plane uOv. Then K_n is a continuous piecewise

[1]In fact, (ii) is valid for any continuous non-decreasing function F.

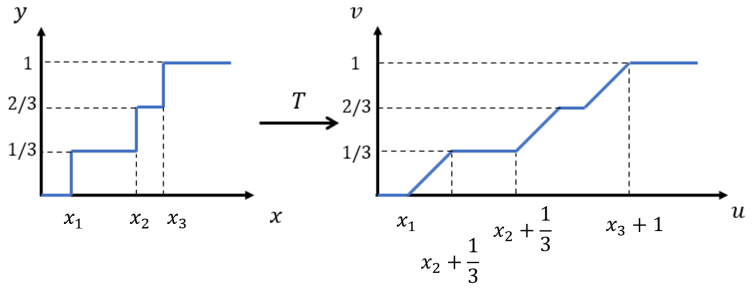


Figure 3.6: Illustration of F_n^c and $T(F_n^c)$ for $\{x_1, x_2, x_3\}$

linear function with vertices $T(A_i)$ $i = 1, 2, \ldots, n$, that has the following form:

$$K_n(u) = \sum_{i=1}^{n} \left\{ \left[u - \left(X_{(i)} + F_n(X_{(i-1)}) \right) \right] \mathbb{1}\left(u \in \left[X_{(i)} + F_n(X_{(i-1)}), X_{(i)} F_n(X_{(i)}) \right] \right) \right\} + \\ + \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\left(X_i + F_n(X_i) \leq u \right), \tag{3.2.19}$$

where $X_{(i)}$, $i = 1, \ldots, n$ are the order statistics of the sample $\boldsymbol{X}_n$ and $F_n(X_{(0)}) \equiv 0$. This is also graphically illustrated in Figure 3.6.

Then the Hausdorff distance $H(F^c, F_n^c)$ given by (3.2.15) is alternatively expressed as follows.

Theorem 3.17. *The one-sample Hausdorff statistic $H(F^c, F_n^c)$ is expressed as*

$$H(F^c, F_n^c) = \sup_t |K_n(t) - K(t)|, \tag{3.2.20}$$

where $K(t), K_n(t)$ are defined in (3.2.18) and (3.2.19), respectively. Furthermore, the supremum is achieved either at the point $t = X_{(i)} + \frac{i-1}{n}$ or at $t = X_{(i)} + \frac{i}{n}$.

Note that the points $(X_{(i)} + \frac{i-1}{n}, K_n(X_{(i)} + \frac{i-1}{n}))$ and $(X_{(i)} + \frac{i}{n}, K_n(X_{(i)} + \frac{i}{n}))$, $i = 1, \ldots, n$, coincide correspondingly with the points $T(A_{2i-1})$ and $T(A_{2i})$, $i = 1, \ldots, n$. Furthermore, obtaining values $\eta(X_{(i)} + \frac{i-1}{n})$ or $\eta(X_{(i)} + \frac{i}{n})$, is equivalent to sequentially solving equations $x + F(x) = X_{(i)} + \frac{i-1}{n}$ or $x + F(x) = X_{(i)} + \frac{i}{n}$. Therefore, computing (3.2.20) is equivalent to computing (3.2.15), using Algorithm 1.

Remark 3.18. Theorem 3.17 is also true when F is discontinuous, with K in (3.2.18) and K_n in (3.2.19) accordingly adjusted so as to determine the parametric equations of $T(F^c)$ and $T(F_n^c)$.

The representation in (3.2.20) allows us to express the cdf as a rectangle probability with respect to the uniform order statistics, following Lemma A.1 in Dimitrova, Kaishev and Tan (2020). The expression is provided in the following theorem.

Theorem 3.19. *When the null cdf $F(x)$ is continuous, the complementary cdf of the one-sample Hausdorff statistic $\mathbb{P}(H(F^c,F_n^c)>q), q\in[0,1]$ can be expressed in terms of the rectangle probability with respect to the order statistics $U_{(i)}, i=1\ldots,n$ of n independent and identically distributed uniform $(0,1)$ random variables U_i, $i=1,\ldots n$, as*

$$\mathbb{P}(H(F^c,F_n^c)>q)=1-\mathbb{P}(a_i\leq U_{(i)}\leq b_i, 1\leq i\leq n), \tag{3.2.21}$$

where

$$\begin{aligned} a_i &= F(F^{-1}(\frac{i}{n}-q)-q),\\ b_i &= F(F^{-1}(\frac{i-1}{n}+q)+q), \end{aligned} \tag{3.2.22}$$

and where $F^{-1}(y)=+\infty$ for $y\geq 1$, $F^{-1}(y)=-\infty$ for $y\leq 0$, $F(+\infty)=1$ and $F(-\infty)=0$.

Expression (3.2.21) can be computed applying the procedure exact-KS-FFT proposed by Dimitrova, Kaishev and Tan (2020) with computational complexity $O(n^2\log(n))$.

Remark 3.20. When $F(x)$ is discrete or mixed, the exact cdf $\mathbb{P}(H(F_n^c,F^c)>q)$ can be expressed similarly as in (3.2.21), with some further adjustments to the values of a_i and b_i.

It should be noted that the cdf $\mathbb{P}(H(F^c,F_n^c)\leq q)$ $q\in[0,1]$ in Theorem 3.19 depends on the null cdf F, since $H(F^c,F_n^c)$ depends on the y and x coordinates. In fact, for any cdf F and $q\in(0,1)$, the sequence $\mathbb{P}(H(F_n^c,F^c)>q)$, $n\to\infty$ converges to 0, as suggested by Corollary 3.12, noting that the KS cdf $\mathbb{P}(\mathcal{D}_n>q)$ converges to 0. Therefore, it is meaningful to consider the asymptotic distribution $\lim_{n\to\infty}\mathbb{P}(\sqrt{n}H(F_n^c,F^c)>\lambda)$ which is done in the following theorem.

Theorem 3.21. *Let $\mathbb{B}_0(t)$, $t\in[0,1]$ be a Brownian bridge with $\mathbb{B}_0(0)=\mathbb{B}_0(1)=0$, $\mathbb{E}[\mathbb{B}_0(t)]=0$, $\mathbb{E}[\mathbb{B}_0(t)\mathbb{B}_0(s)]=s(1-t)$ for $0<s<t<1$. If the null cdf $F(x)$ has a bounded continuous derivative $f(x)$, the asymptotic distribution of the one-sample*

Hausdorff statistic can be expressed as a double-boundary non-crossing probability with respect to the process $\mathbb{B}_0(t)$*. More precisely,*

$$\begin{aligned}&\lim_{n\to\infty}\mathbb{P}(\sqrt{n}H(F_n^c,F^c)>x)\\&=1-\mathbb{P}\Big(-x(1+f(F^{-1}(t)))\le\mathbb{B}_0(t)\le x(1+f(F^{-1}(t))),\forall\ 0\le t\le 1\Big).\end{aligned}\tag{3.2.23}$$

For a particular choice of the null distribution $F(x)$, the analytical expression (3.2.23) can be specified as shown by the following corollary.

Corollary 3.22. *If the null distribution* $F(x)$ *is exponential, i.e.* $F\sim Exp(\alpha)$*, then the asymptotic distribution can be expressed as*

$$\begin{aligned}&\lim_{n\to\infty}\mathbb{P}(\sqrt{n}H(F_n^c,F^c)>x)\\&=1-\mathbb{P}\left(-x(1+\alpha(1-t))\le\mathbb{B}_0(t)\le x(1+\alpha(1-t))\right),\forall\ 0\le t\le 1)\\&=2\sum_{k=1}^{\infty}(-1)^{k-1}e^{-2(1+\alpha)k^2x^2}\end{aligned}\tag{3.2.24}$$

Furthermore, the following theorem establishes an explicit connection between the Kolmogorov-Smirnov and the Hausdorff test statistics and their corresponding (asymptotic) cdfs.

Theorem 3.23. *If the null distribution* $F(x)$ *is uniform, i.e.* $F\sim U(a,a+b)$*, then*

$$H(F^c,F_n^c)=\frac{b}{1+b}\mathcal{D}_n,\tag{3.2.25}$$

and for the cdfs we have

$$\mathbb{P}(\mathcal{D}_n>q)=\mathbb{P}(H(F^c,F_n^c)>\frac{qb}{1+b}).\tag{3.2.26}$$

Moreover,

$$\lim_{n\to\infty}\mathbb{P}(\sqrt{n}H(F^c,F_n^c)\le x)=1-2\sum_{k=1}^{\infty}(-1)^{k-1}e^{-2\frac{(1+b)^2}{b^2}k^2x^2}.\tag{3.2.27}$$

Theorem 3.23 demonstrates the connection between the KS and the Hausdorff statistics and gives the asymptotic distribution of the latter when the null is uniform. Another relation between the H and the KS statistics is established in Theorem

3.36, which indicates that the KS statistic can be viewed as the limit of the scaled H statistic, when the scale parameter goes to infinity.

Remark 3.24. Alexander (1974) has noted the distribution dependence of the Lévy metric in the context of goodness-of-fit testing. Therefore he proposed to use statistic based on Lévy metric to test whether $\{F(X_1), \ldots, F(X_n)\}$ comes from the uniform distribution $U[0,1]$. However, in view of Theorem 3.23, it is not difficult to show that the p-value of Alexander (1974)'s Lévy statistic, will always coincide with that of the KS test as long as the null distribution is continuous.

Let us note that the boundaries in Theorem 3.21 are general and depend on the form of the null distribution. Explicit expressions for the probability that the Brownian bridge $\mathbb{B}_0(t)$ stays within the corridor between two piecewise linear boundaries have been obtained by Novikov et al. (1999) and Pötzelberger and Wang (2001). Note also that an arbitrary null cdf $F(x)$ can be approximated using exponential splines $\hat{F}(x) \in C^1(\mathbb{R})$ that consist of pieces of exponentials $a_i + b_i e^{k_i x}$, $i = 1, \ldots, \kappa$, smoothly joint together at some points called knots, where a_i, b_i and k_i are parameters. It is not difficult to show that such an exponential spline approximation of F would result in piecewise linear boundaries in (3.2.23), and therefore (3.2.23) can be used to compute the asymptotic distributions of the H test for an arbitrary null distribution F, approximated by $\hat{F}(x)$.

3.3 On the Scale Dependence of the H test

It is well known that the majority of the goodness-of-fit tests such as the KS, CvM and AD, and their p-values are location and scale invariant, and this to a great extent motivates their popularity in practice. However, it is not difficult to show that while the proposed H test is location invariant, it is not scale invariant. As a result of this, a major difficulty that arises when one tries to apply the statistic H is that its p-value and hence its power is not invariant to a scale transformation σX, for some constant $\sigma > 0$. This is illustrated in Figure 3.7, where it can be seen that the statistical power of $H(F^c, F_n^c)$ is significantly affected by the choice of the scale coefficient σ. Since power is estimated based on the p-value, Figure 3.7 implies that the p-value of H also changes as σ changes.

To formally illustrate this scale dependence, consider a particular realization

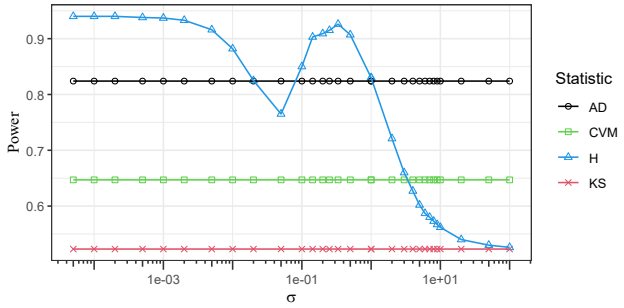


Figure 3.7: Statistical power of AD, CvM, H and KS tests under different scales σ when the sample is from $\mathrm{N}(3/\sigma, \frac{0.2^2}{3\sigma^2})$ and the null distribution is $\mathrm{N}(3/\sigma, \frac{0.2^2}{\sigma^2})$

of the random sample, $\boldsymbol{X}_n(\omega) = \{x_1, \ldots, x_n\}$. Let $\hat{q} = H(F^c, F_n^c(\omega))$ denote the observed value of the H-statistic, with p-value $p = \mathbb{P}(H(F^c, F_n^c) > \hat{q})$. Assume now, that we would like to change the scale and work with the re-scaled sample $\sigma\boldsymbol{X}_n = \{\sigma X_1, \ldots, \sigma X_n\}$ with empirical cdf $F_{n,\sigma}$ and re-scaled hypothetical cdf $F_\sigma(x)$, i.e.

$$F_\sigma(x) = F(x/\sigma), \quad F_{n,\sigma}(x) = F_n(x/\sigma) \equiv \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}(\sigma X_i \leq x), \tag{3.3.1}$$

where σ is appropriately chosen. For example, say $\{x_1, \ldots, x_n\}$ are heights of a sample of n individuals, expressed in meters and we want to work with the same sample of observations but expressed in centimeters, in which case $\sigma = 100$. For the same realization, the observed sample would be $\sigma\boldsymbol{X}_n(\omega) = (\sigma x_1, \ldots, \sigma x_n)$, with empirical cdf $F_{n,\sigma}(\omega) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}(\sigma x_i \leq x)$, observed statistic $\hat{q}'$ and corresponding p-value $\tilde{p} = \mathbb{P}(H(F_\sigma^c, F_{n,\sigma}^c) > \hat{q}')$, where F_σ^c and $F_{n,\sigma}^c$ denote the corresponding planar curves of F_σ and $F_{n,\sigma}$, defined as in (3.2.6). Under the two different scales, the p-values p and $\tilde{p}$ may not coincide.

3.3.1 On the Scale Dependent H Statistic $\mathcal{H}_n(\sigma)$ and Its Related Power

Our aim in this section is to show how the scale dependence property of H can be exploited to optimize its power and therefore eliminate its dependence on σ. To address this aim, we will view the scale coefficient $\sigma > 0$ as a hyperparameter and consider the family of (scaled) Hausdorff statistics

$$\mathcal{H}_n(\sigma) = H(F_\sigma^c, F_{n,\sigma}^c),$$

where F^c_σ and $F^c_{n,\sigma}$ are defined by (3.3.1) and (3.2.6). Given $\sigma > 0$, denote by $L_{n,\sigma}$ the complementary cdf of the statistic $\mathcal{H}_n(\sigma)$ under the null hypothesis, i.e.

$$L_{n,\sigma}(q) = \mathbb{P}(\mathcal{H}_n(\sigma) > q),\ q \in [0,1].$$

Because the power is directly related to the p-value, we will consider the p-value corresponding to a realization $\mathcal{H}_n(\sigma,\omega)$ and define it as a random variable (see Sidak et al., 1999, Chapter 8.3.1, for a similar definition), which we denote as $P_n(\sigma)$, i.e.

$$P_n(\sigma) \equiv P_n(\sigma,\omega) = L_{n,\sigma}(\mathcal{H}_n(\sigma,\omega)) = \mathbb{P}(\mathcal{H}_n(\sigma) > \mathcal{H}_n(\sigma,\omega)), \omega \in \Omega. \tag{3.3.2}$$

Given a realization $\sigma \boldsymbol{X}_n(\omega) = (\sigma x_1, \ldots, \sigma x_n)$, we obtain a specific realization of the p-value $P_n(\sigma,\omega)$. If the sample does not come from the null distribution, we should expect $P_n(\sigma,\omega)$ to be small. The null hypothesis is rejected if the p-value is less than the specified significance level p, i.e., if $P_n(\sigma) \leq p$ and the power of the H test, can be expressed as

$$\pi_n(\sigma) = \mathbb{P}(\overline{H_0}|H_1) = \mathbb{P}(P_n(\sigma) \leq p|H_1) \quad \sigma > 0. \tag{3.3.3}$$

Our aim is to characterize how σ affects the power $\pi_n(\sigma)$. To do so, we use the confidence band approach, previously considered by Mason and Schuenemeyer (1983). For a fixed significance level p and scale parameter σ, the confidence band of the statistic $\mathcal{H}_n(\sigma)$ is defined as the area between two boundaries $\tilde{F}_+(\cdot;\sigma,p)$ and $\tilde{F}_-(\cdot;\sigma,p)$, such that the following events are equivalent

$$\{P_n(\sigma) > p\} = \{\tilde{F}_-(x;\sigma,p) \leq F_n(x) \leq \tilde{F}_+(x;\sigma,p) \text{ for all } x\}, \tag{3.3.4}$$

where $F_n(x)$ is the empirical cdf defined in the original scale. Note that the events on both sides of (3.3.4) depend on σ under the alternative H_1, which is the case we are interested in, to investigate the power $\pi_n(\sigma)$.

Remark 3.25. Let us note that in fact under the null H_0, the probabilities of the

events in (3.3.4) no longer depend on σ, i.e.

$$\mathbb{P}(P_n(\sigma) > p) = \mathbb{P}(\tilde{F}_-(x;\sigma,p) \leq F_n(x) \leq \tilde{F}_+(x;\sigma,p) \text{ for all } x) = 1-p, \tag{3.3.5}$$

which follows from the uniformity of $P_n(\sigma)$.

The following proposition establishes that the boundaries $\tilde{F}_+(\cdot;\sigma,p)$ and $\tilde{F}_-(\cdot;\sigma,p)$ exist and can be found explicitly.

Proposition 3.26. *Given a significance level p and fixed $\sigma > 0$, denote by q^*_σ the $(1-p)$-th quantile of the statistic $\mathcal{H}_n(\sigma)$ under the null, i.e.*

$$q^*_\sigma \equiv q^*(n,\sigma,p) = L^{-1}_{n,\sigma}(p). \tag{3.3.6}$$

Then

$$\begin{aligned} \tilde{F}_-(x;\sigma,p) &= F(x - \frac{q^*_\sigma}{\sigma}) - q^*_\sigma, \\ \tilde{F}_+(x;\sigma,p) &= F(x + \frac{q^*_\sigma}{\sigma}) + q^*_\sigma. \end{aligned} \tag{3.3.7}$$

Note that the boundaries $\tilde{F}_-(x;\sigma,p)$ and $\tilde{F}_+(x;\sigma,p)$ in (3.3.7) hold, regardless of whether F_n is under the null or not and that the confidence band $(\tilde{F}_-(x;\sigma,p), \tilde{F}_+(x;\sigma,p))$ depends on the null distribution F and on the scale parameter σ. The following proposition shows that the power, $\pi_n(\sigma)$, coincides with the probability of $F_n(x)$ exiting the band $(\tilde{F}_-(x;\sigma,p), \tilde{F}_+(x;\sigma,p))$, under the alternative.

Proposition 3.27. *For a fixed $\sigma > 0$, we have*

$$\pi_n(\sigma) = 1 - \mathbb{P}(\tilde{F}_-(x;\sigma,p) \leq F_n(x) \leq \tilde{F}_+(x;\sigma,p) \text{ for all } x | H_1). \tag{3.3.8}$$

Proposition 3.27 suggests that the width and shape of the band, $(\tilde{F}_-(x;\sigma,p), \tilde{F}_+(x;\sigma,p))$, directly affect the power of the statistic $\mathcal{H}_n(\sigma)$, which therefore can be controled by appropriately selecting the scale parameter, σ. In Theorem 3.29, we further characterize the confidence band of $\mathcal{H}_n(\sigma)$ as a function of σ, for a class of null distributions F. Before this, it will be useful to introduce the following auxiliary lemma.

Lemma 3.28. *For any F, $p > 0$ and $\sigma_1 > \sigma_2 > 0$, the following inequalities holds*

$$q^*_{\sigma_2} < q^*_{\sigma_1} < \frac{\sigma_1}{\sigma_2} q^*_{\sigma_2}, \tag{3.3.9}$$

where $q^*_{\sigma_2}$ and $q^*_{\sigma_1}$ are defined as in (3.3.6).

Theorem 3.29. *Let $\iota(x,\sigma) := \tilde{F}_+(x;\sigma,p) - \tilde{F}_-(x;\sigma,p)$, denote the width of the confidence band, for a fixed x. If there exists x_0 such that $\forall\, x > x_0$ the null cdf $F(x)$ is concave, then for any $\sigma_1 > \sigma_2 > 0$, there always exists $x_1 > x_0 \in \mathbb{R}$, such that for any $x > x_1$, $\iota(x,\sigma_1) > \iota(x,\sigma_2)$.*

Corollary 3.30. *If there exists x_0 such that the null distribution $F(x)$ is convex for every $x < x_0$, for any $\sigma_1 > \sigma_2 > 0$, there always exists $x_1 < x_0 \in \mathbb{R}$, such that for any $x < x_1$, $\iota(x,\sigma_1) > \iota(x,\sigma_2)$.*

Corollary 3.31. *If there exists x_0, such that $F(x)$ is convex for $x < x_0$ and $F(x)$ is concave for $x > x_0$ (i.e. F is S-shaped), for any $\sigma_1 > \sigma_2 > 0$, there always exists $x_1 > x_2 \in \mathbb{R}$, such that for any $x \in (-\infty, x_2) \cup (x_1, +\infty)$, $\iota(x,\sigma_1) > \iota(x,\sigma_2)$.*

To conclude, Theorem 3.29 suggests that, if $F(x)$ has a concave right tail, a smaller σ will result in a narrower confidence band at the right tail, which is illustrated in Example 3.32. Therefore, if the null F deviates from the true distribution F_0 mostly in the right tail, then for a fixed significance level p, $F_n(x,\omega)$ would exit the band at the right tail with higher probability. This would make $\mathcal{H}_n(\sigma)$ right-tail sensitive, i.e. $\mathcal{H}_n(\sigma)$ would have higher power when the null and the true distributions tend to differ more in the right tail. To summarize, all these considerations apply to the most commonly used null distributions such as Exponential, Normal, Pareto, Lognormal, Gamma, etc, in which case we achieve an optimal power of the $\mathcal{H}_n(\sigma)$ test by appropriately selecting a small σ. Similarly, following Corollary 3.30, if F has a convex left tail, a smaller σ will make $\mathcal{H}_n(\sigma)$ left tail sensitive, i.e. it would have higher power when the null and the true distributions tend to differ more in the left tail. And following Corollary 3.31 if F is S-shaped, a smaller σ will make $\mathcal{H}_n(\sigma)$ left and right tail sensitive. The case of large σ, is covered in Section 3.3.3.

Example 3.32. Let the null distribution $F \sim Exp(1)$. For $p = 0.5$, the implied confidence bands $(\tilde{F}_-(x;\sigma,p), \tilde{F}_+(x;\sigma,p))$ for different fixed σ, computed based on Proposition 3.26 are shown in Figure 3.8.

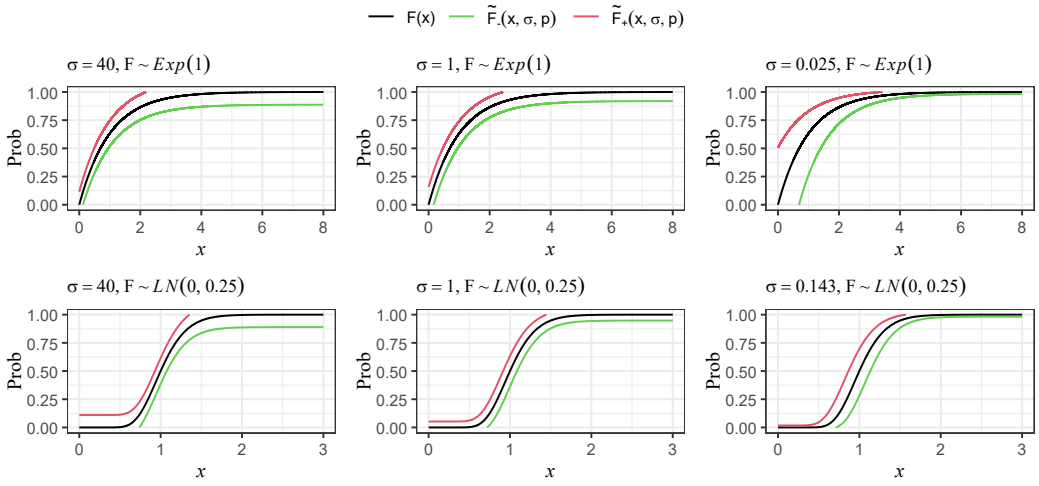


Figure 3.8: Confidence band of $\mathcal{H}_n(\sigma)$ for different σ when the null is $Exp(1)$ with $p = 0.5$ and $n = 50$ or when the null is $LN(0, 0.25^2)$ with $p = 0.2$ and $n = 50$

As can be seen from Figure 3.8, when σ is small, the confidence band is narrower at the right tail (as can be seen from the rightmost panel for $\sigma = 0.025$), whereas when σ is large, the confidence band is narrower at the left tail (as can be seen from the leftmost panel for $\sigma = 40$).

Alternatively, if F is S-shaped, as assumed in Corollary 3.31, a small σ would imply narrower band both at the left and right tails. We illustrate this in Example 3.33 where we choose F to be lognormal.

Example 3.33. Let the null distribution F be Lognormal with location parameter 0 and shape parameter 0.25^2, i.e. $F \sim LN(0, 0.25^2)$. The corresponding confidence bands for different choices of the scale parameter σ are shown in Figure 3.8. As can be seen from the rightmost panel for $\sigma = 0.143$, small σ values lead to bands that are narrow both at the left and right tails. As can be seen from the leftmost panel for $\sigma = 40$, large σ values lead to a narrower band in the body. The middle panel for $\sigma = 1$ suggests that by appropriately selecting σ, one can achieve a more uniformly sized band across the entire support of F. Therefore, the statistic $\mathcal{H}_n(\sigma)$ would have greater power for smaller σ when the alternative and the null cdfs differ either at the left or right tails. And when σ is large, $\mathcal{H}_n(\sigma)$ would be more sensitive in the body, i.e. would have greater power if the alternative and the null distributions are expected to differ more in the body.

In conclusion, the sensitivity of $\mathcal{H}_n(\sigma)$ can be controlled locally over the

support of F by appropriately selecting the scale parameter σ, i.e. accordingly rescaling the sample $\sigma \boldsymbol{X}_n$. In other words, one can select σ so that the band $(\tilde{F}_-(x;\sigma,p), \tilde{F}_+(x;\sigma,p))$ becomes narrower at the area (left and/or right tail or the body) where the null and the true distributions are expected to differ most.

3.3.2 Optimal σ Selection

While Theorem 3.29 and Corollaries 3.30 and 3.31 shed some light on how to change σ, so as to control the bandwidth locally, these results do not suggest a rule for choosing the scale parameter σ, that can be applied in practice. In what follows, we provide such a rule based on insights stemming from these results and some empirical and geometric considerations. One such key insight comes from the ability to decrease σ, and narrow down the confidence band around the tails where F_0 and F are expected to differ most. This motivates us to consider the difference in the left/right tails of F_n (which comes from F_0) and F. Another insight is that the rule for selecting σ should only depend on F and should eliminate the scale dependence of the proposed H statistic.

In order to follow upon these insights, recall Lemma 3.8 which states that the H statistic $H(F^c, F_n^c)$ coincides with the (vertical) side of the largest square that can be inscribed between F^c and F_n^c. Therefore, the σ scaled H statistic $\mathcal{H}_n(\sigma)$ should also coincide with the (vertical) side of the largest square that can be inscribed between the scaled cdfs F_σ^c and $F_{n,\sigma}^c$. Therefore, the value of $\mathcal{H}_n(\sigma)$ will equivalently be found as the vertical side of the largest rectangle among all rectangles with ratio of sides equal to σ, inscribed between F^c and F_n^c, as illustrated in Figure 3.9. These considerations are formally summarized in the following proposition, which is a straightforward restatement of Lemma 3.8 under rescaling with respect to x and is therefore given without proof.

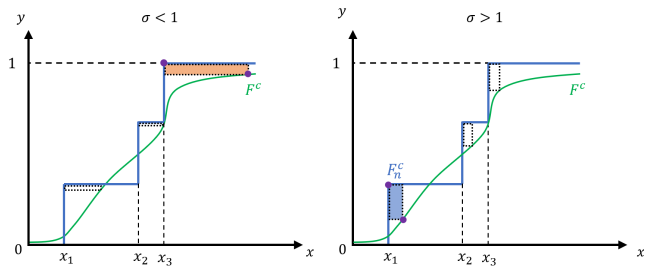


Figure 3.9: $\mathcal{H}_n(\sigma)$ coincides with the vertical side of the colored rectangles for $\sigma < 1$ (left panel) and $\sigma > 1$ (right panel) in the original scale

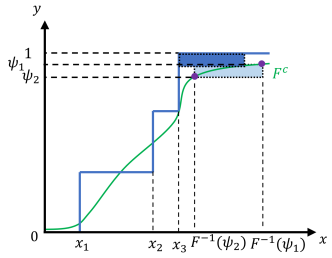


Figure 3.10: The selection of σ for right-tail sensitive statistic following (3.3.11) for F with concave right-tail

Proposition 3.34. *For $\sigma > 0$, if a rectangle $S(P,d;\sigma) = \{(x,y) \in \mathbb{R}^2 : |x - x_P| \leq d/(2\sigma), |y - y_P| \leq d/2\}$, with vertical side d, horizontal side d/σ and center at the point P, can be inserted between the curves F^c and F_n^c, so that $S(P,d;\sigma) \subset \mathcal{G}$, then $\mathcal{H}_n(\sigma) \geq d$. Furthermore,*

$$\mathcal{H}_n(\sigma) = \sup\{d : S(P,d;\sigma) \subset \mathcal{G}\}, \tag{3.3.10}$$

where $\mathcal{G}$ denotes the area between the two planar curves F^c and F_n^c, which has been formally defined in Lemma 3.8.

In summary, the alternative geometric interpretation of $\mathcal{H}_n(\sigma)$ following from Proposition 3.34 is that it coincides with the vertical side of the largest rectangle $S(P,d;\sigma)$ with a ratio of vertical to horizontal sides equal to σ, which can be inscribed in $\mathcal{G}$, i.e. in the area between F^c and F_n^c. This is illustrated in Figure 3.9 for different choices of σ.

On the one hand, when σ becomes smaller, the vertical side d of the largest $S(P,d;\sigma)$ decreases and the horizontal side d/σ increases, i.e. it becomes flatter. Therefore, with σ decreasing, it is less likely that the largest $S(P,d;\sigma)$, i.e. $\mathcal{H}_n(\sigma)$ (cf. (3.3.10)) would occur in the body of $\mathcal{G}$, and more likely it would appear in its tail area (see the left panel in Figure 3.9). On the other hand, as we established, when σ becomes smaller, the confidence band narrows down along the right tail and $\mathcal{H}_n(\sigma)$ becomes more right-tail sensitive. So σ should be chosen so that its corresponding largest $S(P,d;\sigma)$ has a small vertical side and a large horizontal side and appears in the right tail with its top-left vertex coinciding with the right-most vertex of the staircase F_n^c. Unfortunately, this is not formally achievable, since the largest $S(P,d;\sigma)$ depends on the random F_n^c. Instead, we consider a rectangle

defined solely by the null F^c that can be viewed as an approximation to the largest $S(P,d;\sigma)$.

One way to construct such a rectangle and therefore choose σ, is to take its vertical side equal to $\psi_1-\psi_2$, where $\psi_1 \neq \psi_2 \in (0,1)$ are chosen close to each other and close to one, ensuring that the vertical side is small and that the rectangle will appear in the right tail. It is natural to take its horizontal side equal to $F^{-1}(\psi_1)-F^{-1}(\psi_2)$ in order to take into account the shape of the right tail of F (see light blue rectangle in Figure 3.10). This leads to the following choice of σ

$$\sigma^* = \frac{\psi_1-\psi_2}{F^{-1}(\psi_1)-F^{-1}(\psi_2)}, \tag{3.3.11}$$

which can be interpreted as the reciprocal of the quantile density function evaluated at $\psi_1 \approx \psi_2$. The latter plays an important role in exploratory data analysis, extreme value theory, reliability and survival analysis (see Soni et al., 2012).

Alternatively, the values of ψ_1 and ψ_2 can be chosen close to zero, or around 0.5, depending on whether a left-tail-sensitive or body-sensitive H-statistic $\mathcal{H}_n(\sigma)$ is required. Appropriate choices of ψ_1 and ψ_2 are summarized in Table 3.1.[2] In Appendix 3.B, we provide numerical evidence that selecting σ according to (3.3.11), applying Table 3.1 optimizes the power of $\mathcal{H}_n(\sigma)$.

Table 3.1: Suggested choice of ψ_1, ψ_2 in (3.3.11)

Fitting	ψ_1	ψ_2
Left tail	0.05	0.01
Body	0.7 or 0.6	0.4 or 0.3
Right tail	0.99	0.95

Let us note once again that selecting σ according to (3.3.11) requires only the knowledge of the null distribution and therefore can be done prior to testing the goodness-of-fit. The Type I error can be expressed as

$$\mathbb{P}(\overline{H_0}|H_0) = \mathbb{P}(P_n(\sigma) \leq p|H_0) = p,$$

where the last equality follows from (3.3.5) and indicates that the Type I error is a constant and does not depend on σ. However, this is not the case if σ is

[2] In case when the null distribution F is S-shaped, choosing according to either left tail sensitive or right tail sensitive would result in both tail sensitive.

estimated based on the sample, e.g., σ is estimated by the sample range or its standard deviation, in which case the Type I error may increase.

Often in practice, the sample data may be rescaled for different reasons e.g., using different scale coefficients, and presented in different units. It is not difficult to see that under the choice (3.3.11), $\mathcal{H}_n(\sigma^*)$ and its p-value $P_n(\sigma^*)$ will be invariant with respect to the units used to present the data, as stated in the following proposition.

Proposition 3.35. *Let σ^* be computed from (3.3.11) under the initial scale of the null $F(x)$ and the sample $\boldsymbol{X}_n$. Let $F(x)$ and $\boldsymbol{X}_n$ be rescaled using a scale coefficient σ_0, in which case they become $F(x/\sigma_0)$ and $\sigma_0\boldsymbol{X}_n$ respectively. Then the corresponding rescaled σ^* given by (3.3.11) becomes*

$$\sigma_0^* = \frac{\psi_1 - \psi_2}{\sigma_0[F^{-1}(\psi_1) - F^{-1}(\psi_2)]},$$

and the null and the sample rescaled with σ_0^* can be seen to be invariant with respect to σ_0, i.e. $F(x/(\sigma_0\sigma_0^*)) = F(x/\sigma^*)$ and $\sigma_0(\sigma_0^*\boldsymbol{X}_n) = \sigma^*\boldsymbol{X}_n$, and so will be $\mathcal{H}_n(\sigma^*)$ and $P_n(\sigma^*)$.

To conclude, Proposition 3.35 states that the proposed rule of selecting σ using (3.3.11) eliminates the scale dependence of $\mathcal{H}_n(\sigma)$, and leads to the invariance of the p-value $P_n(\sigma)$ under affine transformations of the data.

3.3.3 Limiting Results for $\mathcal{H}_n(\sigma)$ as a function of σ

Another important implication of introducing variability with respect to the scale parameter σ is that as we show in Theorem 3.36 the KS statistic is the limit of the Hausdorff statistic $\mathcal{H}_n(\sigma)$ when $\sigma \to \infty$. This result suggests that, the p-value and power of the Hausdorff test converge correspondingly to the p-value and power of the KS test when $\sigma \to \infty$, as stated in Corollary 3.37.

Theorem 3.36. *For any distribution $F(x)$, we have*

$$\lim_{\sigma\to\infty} \mathcal{H}_n(\sigma,\omega) = \rho_D(F_n(\omega), F) \tag{3.3.12}$$

for every $\omega \in \Omega$.

Corollary 3.37. *Denote by $P_n(\infty)$ and $\pi_n(\infty)$ the p-value and the power of the KS test (which is $\mathcal{H}_n(\infty)$), for any fixed alternative H_1. Then we have*

$$P_n(\infty) = \lim_{\sigma\to\infty} P_n(\sigma), \quad \pi_n(\infty) = \lim_{\sigma\to\infty} \pi_n(\sigma).$$

Let us note that all the results in Section 3.3.1 hold for $\sigma = \infty$. As demonstrated, when σ increases, $\mathcal{H}_n(\sigma)$ becomes more sensitive either in the left tail as in Example 3.32, or in the body as in Example 3.33 or in the right tail (e.g. if F is purely convex). Since the KS test is the limit of $\mathcal{H}_n(\sigma)$ as $\sigma \to \infty$, its power is bounded from above by the power of the KS test, if the null deviates from the true at either of these areas.

Theorem 3.38. *For any distribution $F(x)$, we have*

$$\lim_{\sigma\downarrow 0} \frac{1}{\sigma}\mathcal{H}_n(\sigma,\omega) = \sup_{0<t<1} |F_n^{-1}(t,\omega) - F^{-1}(t)| \tag{3.3.13}$$

for every $\omega \in \Omega$, where F_n^{-1} is the inverse of F_n. If F has unbounded support, the limit in (3.3.13) diverges.

As a consequence of Theorem 3.38, if the null F has unbounded support, $\frac{1}{\sigma}q_\sigma^* \to \infty$ as $\sigma \downarrow 0$, where q_σ^* is defined in (3.3.6). Therefore, for the boundaries (3.3.7), we have $\lim_{\sigma\downarrow 0} \tilde{F}_-(x;\sigma,p) \equiv q_\sigma^*$ and $\lim_{\sigma\downarrow 0} \tilde{F}_+(x;\sigma,p) \equiv 1 - q_\sigma^*$.

3.4 Asymptotic Power

In Section 3.3, we showed that when the sample size n is finite, the power of the proposed H statistic can be optimized by appropriately selecting the scale parameter σ (according to (3.3.11)), so that it significantly exceeds the power of the major classical tests, including the KS test, as also demonstrated in Section 3.5.

In this section, we go a step further and provide some asymptotic results for the power, $\pi_n(\sigma)$ of the H test, when $n \to \infty$, expressing it as a boundary crossing probability and utilizing the concept of exact Bahadur slope and Bahadur efficiency. The importance of these results are illustrated in Example 3.43, where we show that the proposed H statistic is better suited to test tail differences than the KS statistic not only when n is finite, but also asymptotically when $n \to \infty$. The importance of studying the asymptotic power of goodness-of-fit tests has been highlighted in the

seminal work by Janssen (2000).

Let us assume that a sequence of i.i.d. random variables $\{X_1,\ldots,X_n,\ldots\}$ come from an unknown true distribution F_0. For our asymptotic considerations we will keep the null hypothesis $H_0 : F_0 = F$, but will make the alternative more specific, i.e. H_1: $F_0(x) = G(x;n)$, where $G(x;n)$ is indexed by n. Here, we will be interested in the behaviour of $\pi_n(\sigma)$ and $\pi_n(\infty)$ as n approaches infinity. To characterise these asymptotic powers, we could take two approaches. First is to consider describing the speed of convergence of the p-values $P_n(\sigma)$ and $P_n(\infty)$ for a fixed alternative, i.e. $G(x;n) \equiv G(x)$, $n = 1,2,\ldots$, leading to the study of the Bahadur efficiency of $\mathcal{H}_n(\sigma)$. The concept of the Bahadur relative efficiency is illustrated in Example 3.43, Section 3.4.2.

Second is to choose an appropriate contiguous alternative, resulting in the study of the asymptotic power $\pi_n(\sigma)$ and $\pi_n(\infty)$, $n \to \infty$. Note that the second approach cannot be applied if a fixed alternative is chosen, since both $\pi_n(\sigma)$ and $\pi_n(\infty)$ would converge to 1. Theorem 3.39 gives a representation of the asymptotic power of the H test as the boundary crossing probability of a shifted Brownian bridge.

3.4.1 Asymptotic Power under A Contiguous Alternative

As mentioned, in order to study the asymptotic behavior of $\pi_n(\sigma)$ and $\pi_n(\infty)$, $n \to \infty$, we require a contiguous alternative, i.e. we assume that the alternative $G(x;n)$ converges to the null at the rate $\frac{1}{\sqrt{n}}$. Therefore, we assume that $G(x;n)$ has the inverse

$$G^{-1}(t;n) = F^{-1}(t) + \frac{1}{\sqrt{n}}\delta(t),\ t \in [0,1), \tag{3.4.1}$$

where $\delta(t)$ is a bounded continuous function defined on $(0,1)$, characterizing how the alternative differs from the null F. Since we are interested in the limiting behaviour of $\pi_n(\sigma)$, the required $\delta(t)$ ensures $G(x;n)$ are cdfs for sufficiently large n. It is well known (see Milbrodt and Strasser, 1990) that the asymptotic power of the KS test $\pi_n(\infty)$ is related to the shifted Brownian bridge. This result can be generalized to the asymptotic power $\pi_n(\sigma)$ of the Hausdorff statistic, as shown in the following theorem.

Theorem 3.39. *Consider the null hypothesis $H_0 : F_0 = F$ with density f, and assume that the sequence $\{X_1,\ldots,X_n,\ldots\}$ comes from $G(x;n)$ with inverse defined as*

in (3.4.1). For fixed $\sigma > 0$ and signficance level p, the asymptotic power of the scaled Hausdorff statistic $\mathcal{H}_n(\sigma)$ can be expressed as

$$\lim_{n\to\infty} \pi_n(\sigma) = \mathbb{P}(\gamma_\sigma(t) \le \mathbb{B}_0(t) - \zeta(t) \le \gamma_\sigma(t), \forall\ 0 \le t \le 1), \tag{3.4.2}$$

where $\mathbb{B}_0(t)$ is a Brownian Bridge defined as in Theorem 3.21,

$$\zeta(t) = \delta(t) f(F^{-1}(t)),\ \gamma_\sigma(t) = \lambda_p(\sigma)(1 + \frac{1}{\sigma} f(F^{-1}(t))),$$

and where $\lambda_p(\sigma)$ is the quantile of the asymptotic distribution of $\mathcal{H}_n(\sigma)$, i.e. it is the solution of

$$p = \mathbb{P}\left(|\mathbb{B}_0(t)| \le \lambda_p(\sigma)\left(1 + \frac{1}{\sigma} f(F^{-1}(t))\right), \forall\ 0 \le t \le 1\right).$$

Corollary 3.40. *The asymptotic power of the KS statistic is*

$$\lim_{n\to\infty} \pi_n(\infty) = \mathbb{P}(|\mathbb{B}_0(t) - \zeta(t)| \le \lambda_p(\infty), \forall\ 0 \le t \le 1),$$

where $\lambda_p(\infty)$ is the quantile of the asymptotic distribution of the KS statistic, which is also the solution of $p = \mathbb{P}(|\mathbb{B}_0(t)| \le \lambda_p(\infty), \forall\ 0 \le t \le 1)$.

Similar result for the KS statistic is also obtained in Milbrodt and Strasser (1990).

3.4.2 The Bahadur Relative Efficiency

The p-value of many well-known goodness-of-fit statistics among which KS, CvM, and AD converge to zero exponentially, i.e. $O(e^{-cn})$ when the null is $H_0 : F_0(x) = F(x)$ and the sample $\{X_1, \ldots, X_n\}$ comes from a fixed alternative with cdf $G(x; n) \equiv G(x)$. Therefore, Bahadur (1971) introduced the Bahadur exact slope c as a characterization for the speed of the exponential convergence. Let us note that the existence of the slope c implies an exponential rate of convergence. We adopt the definition from Bahadur (1971) and introduce the corresponding Bahadur exact slope of $\mathcal{H}_n(\sigma)$ as

$$c(\sigma) = -\lim_{n\to\infty} 2n^{-1} \log P_n(\sigma) \quad a.s., \tag{3.4.3}$$

which can be interpreted as the speed of exponential convergence of the p-value $P_n(\sigma)$. For consistency, denote $c(\infty) = \lim_{\sigma\to\infty} c(\sigma)$ the Bahadur exact slope of the KS statistic. As known from the literature (see e.g. Abrahamson, 1967), the exact slope $c(\infty)$ of the KS test exists. Furthermore, as shown in Theorem 3.41, the limit in (3.4.3) exists for any $\sigma > 0$ and therefore $c(\sigma)$ is well defined.

It is worth pointing out that for any goodness-of-fit statistic the corresponding exact slope depends on the choice of the null F and the alternative G and is bounded from above by $2\rho_{KL}(G,F)$, where ρ_{KL} denotes the Kullback–Leibler information measure, defined as

$$\rho_{KL}(G,F) = \int_{-\infty}^{+\infty} g(x) \log \frac{g(x)}{f(x)} dx, \tag{3.4.4}$$

where $f(x)$ and $g(x)$ are the density functions of $F(x)$ and $G(x)$. Thus, $c(\sigma) \leq 2\rho_{KL}(G,F)$. Furthermore, this bound is attained by a class of statistics with certain regularity conditions (see e.g. Nikitin, 1995). The quantity

$$\frac{c(\sigma)}{2\rho_{KL}} \leq 1 \tag{3.4.5}$$

is also referred to in the literature as the absolute Bahadur efficiency.

Given the sequence $\{X_1,\ldots,X_n,\ldots\}$ and the null distribution $F(x)$, under the alternative hypothesis $G(x)$, let $N(\sigma_i,p)$, be the smallest sample size required for the null to be rejected using the $\mathcal{H}_n(\sigma_i)$ statistic for fixed σ_i, $i = 1,2$ at a significance level p. Then we consider,

$$\lim_{p\to 0} \frac{N(\sigma_1,p)}{N(\sigma_2,p)} = \frac{c(\sigma_2)}{c(\sigma_1)} \quad a.s., \tag{3.4.6}$$

which is known as the Bahadur relative efficiency (see Bahadur, 1971, page 27 therein). As noted by Bahadur, comparing the exact slopes of two goodness-of-fit statistics leads to comparing their relative efficiencies. Therefore, in order to compare the powers of the KS and $\mathcal{H}_n(\sigma)$ statistics, we need to estimate their corresponding exact slopes $c(\infty)$ and $c(\sigma)$. While $c(\infty)$ is known for the assumed F and G (see e.g. Abrahamson 1967), estimating the exact slope $c(\sigma)$ can be done by applying the following theorem.

Theorem 3.41. *Under the alternative $\rho_\infty : F_0 = G$, the exact slope defined in (3.4.3)*

exists, and can be expressed as $c(\sigma) = \frac{1}{2} f_\sigma(H(F_\sigma, G_\sigma))$*, where* F_σ *and* G_σ *denote the null and the alternative scaled by* σ*,*

$$f_\sigma(x) = \inf_{0<t<1} \min\{r(t, x + F(F^{-1}(t) + \frac{x}{\sigma}) - t), r(1-t, x - F(F^{-1}(t) - \frac{x}{\sigma}) + t)\}, \tag{3.4.7}$$

and where

$$r(x,y) = \begin{cases} (x+y)\log(\frac{x+y}{x}) + (1-x-y)\log(\frac{1-x-y}{1-x}), & 0 < x < 1-y \\ +\infty, & 1-y \le x < 1; \end{cases}$$

Corollary 3.42. *When the null distribution* F *is continuous, for sufficiently small* x*,* $f_\sigma(x)$ *can be represented as*

$$f_\sigma(x) = x^2 \left[\inf_{0<t<1} \frac{\left(1 + \frac{1}{\sigma} f(F^{-1}(t)\right)^2}{t(1-t)} \right] + O(x^3). \tag{3.4.8}$$

In the following example, we illustrate how the Bahadur relative efficiency can be used to compare the asymptotic powers of the KS and the scaled $\mathcal{H}_n(\sigma)$ tests for $\sigma = \sigma^*$ selected according to (3.3.11) and assuming that the alternative differs from the null only in the tail, which is achieved by an appropriate splicing construction for the alternative.

Example 3.43 (A Splicing Construction of G)**.** We take the alternative G to be a spliced distribution with support $[\underline{x}, \infty)$ and density $g(x)$, that coincides with the null F (with density f) up to a splicing point $\underline{x} < C < \infty$, after which it coincides with a distribution $G_0(x; \theta, \alpha)$ (with density $g_0(x; \theta, \alpha)$) depending on shape and scale parameters θ, α. The density $g(x)$ can then be expressed as

$$g(x) = \begin{cases} \phi_1 g_1(x), & \underline{x} \le x \le C \\ \phi_2 g_2(x), & x > C \end{cases}, \tag{3.4.9}$$

where $\phi_1, \phi_2 > 0$ are the splicing weights, satisfying $\phi_1 + \phi_2 = 1$, $g_1(x)$ and $g_2(x)$ are proper probability density functions defined on $[\underline{x}, C]$ and $[C, \infty)$ respectively. We also require the following conditions to be met.

1) $C = F^{-1}(\phi_1)$,

2) $g(x) = f(x)$, for $x \in [0, C]$.

3) $g(x) = g_0(x; \theta, \alpha)$ for $x > C$ and as a consequence $G_0(C; \theta, \alpha) = \phi_1$, where α and θ are appropriately chosen.

It is not difficult to verify that conditions 1) - 3) ensure that $g_1(x) = f(x)/\phi_1$ and $g_2(x) = g_0(x; \theta, \alpha)/\phi_2$ intergrate to one, and that the alternative G has the body of the null F and the tail of G_0. In order to illustrate the construction, we take F to be exponential $F \sim Exp(1)$, $\phi_1 = 0.8$, $\underline{x} = 0$, $C = -\log(0.2)$ which follows from condition 1), and represent the tail by $G_0(x; \theta, \alpha) \equiv G_0(x/\alpha; \theta, 1)$ defined by θ and α. Clearly, if one fixes α, the value of θ is uniquely determined from $G_0(C; \theta, \alpha) = \phi_1$ of condition 3). Therefore, G is fully controlled by the choice of α. To determine α, we also require G to be as close to F as possible, in terms of the Kullback–Leibler information measure $\rho_{KL}(G, F)$, (c.f (3.4.4)) which is specified in the next condition.

4) (a) If $\inf_\alpha \rho_{KL}(G, F) \neq 0$, we choose $\alpha^* = \arg\inf_\alpha \rho_{KL}(G, F)$.

   (b) If $\inf_\alpha \rho_{KL}(G, F) = 0$, then we choose α^* such that $\rho_{KL}(G, F) = 0.001$. When the latter equality does not ensure that α^* is unique, we further require that $F(x) \leq G(x)$ for $x > C$ to guarantee its uniqueness.

In summary, given the class G_0, we are able to find unique α^* and θ^* that characterize G so that conditions 1) - 4) are met. In Table 3.2 we present the Bahadur absolute efficiencies $\frac{c(\sigma^*)}{2\rho_{KL}}$, $\frac{c(\infty)}{2\rho_{KL}}$, for the H and the KS tests and the Bahadur relative efficiency $\frac{c(\sigma^*)}{c(\infty)}$, for the H versus the KS test, for different tails G_0 of G. For all tail distributions G_0, $\inf_\alpha \rho_{KL} = 0$. These efficiencies are eveluated following (3.4.5) and (3.4.6), where $c(\cdot)$ and ρ_{KL} are computed based on Theorem 3.41 and Eq. (3.4.4). As can be seen from the last column of Table 3.2, under the different choices for the tail part G_0 of G, the $\mathcal{H}_n(\sigma^*)$ test is between 6 and 267 times more efficient than the KS test, in the sense of the Bahadur relative efficiency. The scale parameter σ^* in $\mathcal{H}_n(\sigma^*)$ is chosen according to (3.3.11) with $(\psi_1, \psi_2) = (0.99, 0.95)$ (cf. Table 3.1) and $F \sim Exp(1)$. According to (3.4.6), when $p \to 0$, the sample size $N(\sigma^*, p)$ required to reject the null using $\mathcal{H}_n(\sigma^*)$ is asymptotically between 6 and 267 times less than the sample size $N(\infty, p)$ required using the KS, if the null and alternative differ in the tail. In other words, for a given sufficiently large sample size n, the power of $\mathcal{H}_n(\sigma^*)$ is much higher than the power of KS.

Table 3.2: Absolute and Relative Bahadur Efficiency of the KS test and the scaled $\mathcal{H}_n(\sigma)$ test with $\sigma = \sigma^*$ following (3.3.11) with $\psi_1 = 0.99$ and $\psi_2 = 0.95$ for different tail specifications G_0 of the alternative.

	$G_0(x;\theta,\alpha)$	ρ_{KL}	θ^*	α^*	$\frac{c(\sigma^*)}{2\rho_{KL}}$	$\frac{c(\infty)}{2\rho_{KL}}$	$\frac{c(\sigma^*)}{c(\infty)}$
Lognormal tail	$\Phi(\frac{\ln(x)-\alpha}{\theta})$	0.0024	-0.1314	0.7216	0.013	0.0004	32.39
Gumbel tail	$e^{-e^{-(x-\theta)/\alpha}}$	0.00011	0.1887	0.9472	0.035	0.0018	19.84
Fréchet tail	$e^{-(x/\alpha)^\theta}$	0.0150	2.9353	0.9655	1.32e-3	4.9e-6	267.4
Weibull tail *	$(1-e^{-(x/\alpha)^\theta})\mathbb{1}(x \geq 0)$	0.0010	0.9321	0.9659	0.158	0.0220	7.17
Pareto tail *	$(1-(\frac{\alpha}{x+\alpha})^\theta)\mathbb{1}(x \geq 0)$	0.0010	21.114	20.319	0.118	0.0117	10.1
Gamma tail *	$\int_0^x \frac{1}{\Gamma(\theta)\alpha^\theta} t^{\theta-1}e^{-t/\alpha}dt$	0.0010	0.8528	1.1592	0.142	0.0228	6.25

It would also have been interesting to compare the asymptotic power of the H test with that of other classical tests, e.g., the CvM and AD tests, using the Bahadur relative efficiency, (3.4.6). Although expressions for their exact slopes, needed in (3.4.6), have been obtained by Nikitin (1995), they are difficult to compute numerically for given F and G. For this reason, we have not included such comparisons in Table 3.2.

3.5 Numerical Studies on the Use of $\mathcal{H}_n(\sigma)$ for Fitting Tails

We showed in Sections 3.3 and 3.4, (see Proposition 3.27, Theorems 3.29, 3.41 and Table 3.2) that as σ decreases, $\mathcal{H}_n(\sigma)$ becomes more sensitive in capturing tail deviations. This is illustrated and supported, based on the simulated Example 3.44 of Sections 3.5.1 and the real data Example 3.45 in Section 3.5.2. Example 3.44 shows that when the null and alternative differ in the right tail, the power of the scaled $\mathcal{H}_n(\sigma^*)$ test with σ^* selected according to (3.3.11), is significantly higher than that of other classical goodness-of-fit tests. In Example 3.45, we first specify the null cdf F as a spliced distribution from the Mixed Erlang-Pareto (ME-P) family considered by Reynkens et al. (2017) as a global, body and tail fitting model. We then fit the latter to an appropriately selected subset of a Lloyd's aviation loss data set and then use the remaining subset to test the null $H_0 : F_0 = F$, applying the KS, CvM, AD and $\mathcal{H}_n(\sigma^*)$ tests. By analyzing the PP and QQ plots of Fig 3.12 and the goodness-of-fit testing results summarized in Table 3.4, we show that, among all tests, the $\mathcal{H}_n(\sigma^*)$ test is better suited for detecting the tail difference between the sample and the null.

3.5.1 Simulated Example

This section provides a simulated example to illustrate the power of $\mathcal{H}_n(\sigma)$ when the null and alternative differ in the tail following the results of Section 3.3. We show that if the sample comes from a distribution whose tail differs from that of the null, $\mathcal{H}_n(\sigma^*)$ has a higher power than KS, CvM and AD, when σ^* is chosen according to (3.3.11).

Example 3.44. We apply the splicing construction as in Example 3.43, where the alternative G has a density $g(x)$ defined by (3.4.9), satisfying conditions 1)-3). In other words, G coincides with F in the body and has the tail of G_0. We assume that both F and G_0 come from the family of (unshifted) Fréchet distributions, which are characterized by scale and shape parameters α and θ. More precisely, the cdf of a Fréchet distribution $\Phi(x;\alpha,\theta)$ is given as

$$\Phi(x;\alpha,\theta) = \begin{cases} \exp\Big(-(1+\frac{\theta x}{\alpha})^{-\frac{1}{\theta}}\Big) & x > -\frac{\alpha}{\theta} \\ 0 & x \leq -\frac{\alpha}{\theta} \end{cases}. \tag{3.5.1}$$

In order to compute and compare the powers of $\mathcal{H}_n(\sigma)$, KS, CvM and AD, we take $F \sim \Phi(x;0.3,0.3)$, $\underline{x} = -1$, $C = (-\log 0.8)^{-0.3} - 1$, $F(C) = \phi_1 = 0.8$, and $G_0 \sim \Phi(x;\theta_0,\alpha_0)$. As in Example 3.43, condition 3) ensures the uniqueness of α_0 for a fixed shape parameter θ_0. Therefore, the choice of θ_0 fully determines the shape of G. In this example, we are interested in comparing numerically the powers of KS, CvM and AD with that of the $\mathcal{H}_n(\sigma)$ statistic, under different tail alternatives from the Fréchet family (3.5.1). For this reason, we choose $\theta_0 = 0.3 + \Delta$, where $\Delta = 0.2, 0.4, 0.6, \ldots, 3.2$ reflects the tail difference between F and G. As shown in the left panel of Figure 3.11, with Δ increasing, the difference between G and F in the tail also increases.

Then a sample $\boldsymbol{X}_n$ with $n = 50$ is simulated from $G(x)$, and the powers of KS, Cramér–von Mises (CvM), Anderson-Darling (AD) and $\mathcal{H}_n(\sigma^*)$ tests are estimated by their corresponding frequencies of rejection of the null $H_0 : F_0 = F$ at significance level $p = 0.05$. To compute the p-value of CvM and AD tests, we use the R package **goftest** (Faraway, Marsaglia, Marsaglia and Baddeley, 2021). To compute the p-value of KS and the $\mathcal{H}_n(\sigma^*)$ tests, we use the exact-KS-FFT procedure from the R

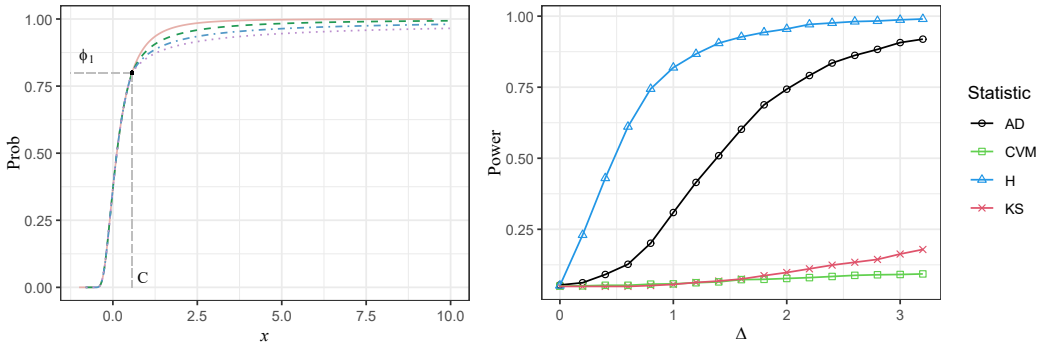


Figure 3.11: Left Panel: the null (solid) $F(x) \sim \Phi(x; 0.3, 0.3)$ and the alternative G with tail $G_0(x) \sim \Phi(\theta_0, \alpha_0)$, where the splicing point $C = (-\log 0.8)^{-0.3} - 1$, $F(C) = \phi_1 = 0.8$ and $\theta_0 = 0.3 + \Delta$ for $\Delta = 0.4$ (dashed), 0.8 (dotdash), 1.2 (dotted); Right Panel: the powers of AD, CvM, $\mathcal{H}_n(\sigma^*)$ and KS tests, as functions of Δ.

package **KSgeneral** (Dimitrova et al., 2020) to implement (3.2.21) in Theorem 3.19. Since we are interested in fitting the right tail, we choose $\sigma = \sigma^*$ defined in (3.3.11) with $\psi_1 = 0.99$, $\psi_2 = 0.95$ (cf. Table 3.1). The right panel of Figure 3.11 shows the powers of the $\mathcal{H}_n(\sigma^*)$, KS, CvM and AD tests, as functions of Δ.

As can be seen, the power of the proposed H statistic is significantly higher than the powers of the classical tests, AD, CvM, and KS. This illustrates very well the advantage of the H statistic in examining tail differences. This is also supported by the result of Example 3.43, which shows that the relative efficiency of H to KS is significantly higher.

3.5.2 Real Data Example

For this example, we use real data from a Llody's syndicate. The dataset comprises 1,743 aviation insurance losses which have occurred within the period from 2014 to 2019. In aviation insurance, financial losses tend to exhibit a heavy right tail with low frequency but high severity insurance claims, as is the case with our dataset (see the right panel of Figure 3.13). Accurately modeling the tail behavior of such losses, i.e. quantifying the likelihood and magnitude of high-consequence events, is critical for robust risk assessment.

Our purpose is to demonstrate the advantage of applying the proposed $\mathcal{H}_n(\sigma)$ statistic for testing tail difference compared to the KS, AD and CvM statistics. Note that in goodness-of-fit testing, the null cdf, F must be fully specified a priori, with no parameters estimated based on the tested sample. To ensure this, we randomly

divide the dataset into two groups, Groups A and B, based on whether the loss occurred on an odd-numbered (Group A) or even-numbered (Group B) day of the month. This splitting criterion is random, as the loss date is independent of the loss amount. We then use Group A to fit the null cdf, F belonging to a class of distributions and subsequently test whether Group B comes from this fitted null. To further demonstrate the randomness of the split, we present summary statistics

Table 3.3: Summary statistics for the aviation data Group A (odd) and Group B (even) losses that have occurred correspondingly on the odd or even days of the month

	# Obs.	Mean	St.Dev.	Min	Q1	Q2	Q3	Max
Group A (Odd)	930	1,030,282	4,363,370	0.5	6,257	25,000	209,618	63,615,000
Group B (Even)	813	923,381	5,201,619	19.3	5,589	26,180	210,712	116,936,064
Total	1,743	980,420	4,771,612	0.5	5,771	25,304	210,349	116,936,064

of Groups A and B in Table 3.3. As shown, the summary statistics for both groups are very similar, indicating that the two samples are similar in distribution. Therefore, we expect that any differences identified by the goodness-of-fit test will not be attributable to differences between the groups, but rather to deviations from the fitted null. Details of how this is implemented are summarized as follows.

Example 3.45. We first asume that the null is a Mixed-Erlang and Pareto (ME-P) spliced distribution which we fit to the aviation losses dataset. The latter spliced distribution has been suggested by Reynkens et al. (2017), as a model that can successfully capture small to medium losses in the body and large losses in the tail, referring to this as a "global fit" strategy. A ME-P spliced distribution is defined by the following density

$$f_{ME-P}(x) = \begin{cases} \phi_1 \frac{f_{ME}(x;\boldsymbol{\alpha},\boldsymbol{r},\theta)}{F_{ME}(C_1;\boldsymbol{\alpha},\boldsymbol{r},\theta)} & 0 \leq x < C_1 \\ \phi_2 f_P(x;\theta_p,C_1,\infty) & x \geq C_1 \end{cases}, \tag{3.5.2}$$

where ϕ_1 and ϕ_2 are the splicing weights with $\phi_1, \phi_2 > 0$ and $\phi_1 + \phi_2 = 1$, C_1 is the splicing point, f_{ME} and F_{ME} are the density and cdf of the Mixed Erlang distribution, and f_P is the density of the truncated Pareto distribution. The latter

density functions are defined as

$$f_{ME}(x;\boldsymbol{\beta},\boldsymbol{r},\alpha)=\sum_{j=1}^{M}\beta_j\frac{\exp(-x/\alpha)x^{r_j-1}}{\alpha^{r_j}(r_j-1)!},\ x>0,$$

$$f_P(x;\theta_p,C_1,C_2)=\frac{\theta_p C_1^{\theta_p}}{x^{\theta_p+1}\left(1-(\frac{C_1}{C_2})^{\theta_p}\right)},\ C_1\le x<C_2,$$

where M is the number of Erlang distributions used for the mixture, $\boldsymbol{r}=(r_1,\ldots,r_M)$ are their shape parameters, satisfying $r_1<r_2<\cdots<r_M$, $r_i=1,2,3,\ldots$, α is their common scale parameter and $\boldsymbol{\beta}=(\beta_1,\ldots,\beta_M)$ are the mixing weights, i.e. $\beta_i>0$ $i=1,\ldots,M$, and $\sum_{i=1}^{M}\beta_i=1$. For the truncated Pareto distribution $f_P(x;\theta_p,C_1,C_2)$, it is required that $0<C_1<C_2$, $\theta_p>0$, where C_2 is allowed to take infinity. In the ME-P spliced distribution, we let $C_2=\infty$. Given a data sample and a chosen splicing point C_1, Reynkens et al. (2017) estimate $\boldsymbol{r}$, $\boldsymbol{\beta}$, α and θ_p via maximum likelihood, and use the Bayesian Information Criterion (BIC) to select the parameter M, which discourages over-fitting. Estimation of the latter parameter can be performed using the R package **ReIns** (Albrecher et al., 2017), which we also use in this Example.

Then we apply the procedure mentioned above to fit the ME-P spliced distribution (3.5.2) to the data in Group A and consider it as our null F. Note that the estimated M will depend on the choice of the splicing points C_1. Therefore, we further estimate the pair C_1 and M corresponding to the minimum BIC. All the estimated parameters are shown in Table 3.4.

Table 3.4: Parameters of the ME-P null fitted to Group A data for C_1 equal to the 64.95% data quantile and the results of the goodness-of-fit testing based on Group B data

Estimated Parameters of the ME-P null for Group A		Goodness of fit for Group B		
Splicing Point	$C_1=91,464.41$	**Statistics**	***p*-values**	**Values**
	$(\phi_1,\phi_2)=(0.6495,\ 0.3505)$	KS	0.3711	0.0321
Mixed Erlang (Body)	$\beta=(0.8325797,0.1674203)$	CvM	0.3581	0.1609
$(M=2)$	$r=(1,6)$	AD	0.2443	1.2643
	$\alpha=10605.33$	$\mathcal{H}_n(\sigma^*)$	0.001837	0.0116
Pareto Tail	$\theta_p=0.4968537$	with $\boldsymbol{\psi}=(0.99,0.95)$		

Then we use the KS, AD, CvM, and the proposed $\mathcal{H}_n(\sigma)$ statistics to test whether the estimated ME-P null F fits well the data in Group B, i.e. to test the null $H_0:F_0=F$. Since our focus is testing the goodness-of-fit in the tail, we select the scale parameter σ for $\mathcal{H}_n(\sigma)$, following (3.3.11) with $\boldsymbol{\psi}=(\psi_1,\psi_2)=(0.99,0.95)$

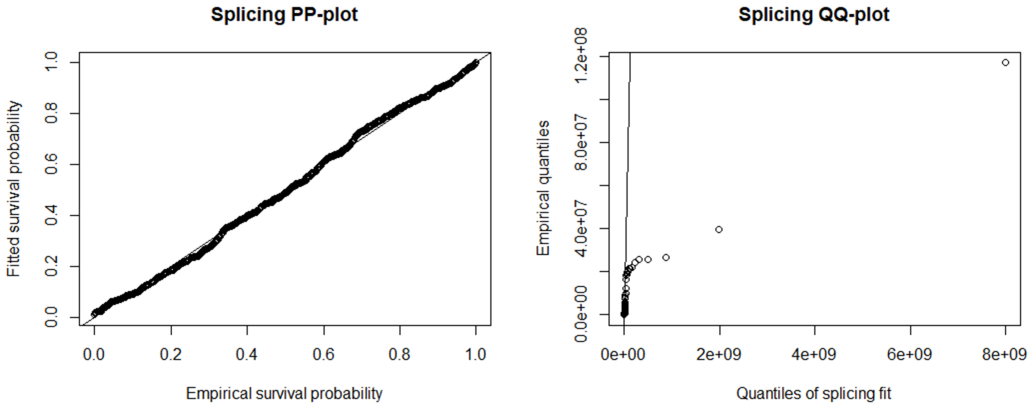


Figure 3.12: The PP and QQ plots for the ME-P distribution in Table 3.4

(cf. Table 3.1), based on the fitted ME-P null distribution F, which yields $\sigma^* = 3.541102 \times 10^{-10}$. The values of these statistics and their p-values are summarized in Table 3.4.

As can be seen from Table 3.4, for a significance level $p = 0.05$, the null H_0 is rejected applying the $\mathcal{H}_n(\sigma^*)$ test, whereas it is not rejected by the KS, CvM, and AD tests. This suggests that the null F does not fit well the right tail of the Group B data. This is also illustrated by the QQ plot in the right panel of Figure 3.12. As can be seen, there is a perfect alignment with the 45-degree diagonal on the PP plot in the left panel, whereas the QQ plot clearly indicates a significant tail deviation. Let us note once again that among all competing tests, the scaled Hausdorff test $\mathcal{H}_n(\sigma^*)$ is the only one that captures this deviation.

In order to remedy this tail discrepancy, we further enhance the ME-P null distribution by introducing a second splicing point C_2 and consider a Mixed-Erlang, Pareto and Pareto (ME-P-P) spliced distribution, with density

$$f_{ME-P-P}(x) = \begin{cases} \phi_1 \frac{f_{ME}(x;\boldsymbol{\alpha},\boldsymbol{r},\theta)}{F_{ME}(C_1;\boldsymbol{\alpha},\boldsymbol{r},\theta)} & 0 < x < C_1 \\ \phi_2 f_P(x;\theta_{p_1},C_1,C_2) & C_1 < x < C_2\,, \\ \phi_3 f_P(x;\theta_{p_2},C_2,\infty) & x > C_2 \end{cases} \tag{3.5.3}$$

where $\phi_1, \phi_2, \phi_3 > 0$, $\phi_1 + \phi_2 + \phi_3 = 1$. In this setting, the choice C_1 is the same as in the ME-P null distribution. To select C_2, we use the mean excess plot criterion instead and thus select C_2 to be the 99-percentile of the data in Group A. Given

C_1 and C_2, the same procedure as for the ME-P null has been used to estimate the parameters $\boldsymbol{\alpha}, \boldsymbol{r}, \boldsymbol{\phi}, \theta, \theta_{p_1}$ and θ_{p_2} in (3.5.3), and the results are summarizes in Table 3.5.

Table 3.5: Parameters of the ME-P-P null fitted to Group A data for C_1 and C_2 equal to the 64.95% and 99% data quantiles and the results of the goodness-of-fit testing based on Group B data

Estimated Parameters of the ME-P-P null for Group A		Goodness of fit for Group B		
Splicing Point	$C_1 = 91,464.41$, $C_2 = 25,142,116.72$	**Statistics**	***p*-values**	**Values**
	$(\phi_1, \phi_2, \phi_3) = (0.6494, 0.3398, 0.0108)$	KS	0.1655	0.0391
Mixed Erlang	$\beta = (0.8325797, 0.1674203)$	CvM	0.3380	0.1685
$(M = 2)$	$r = (1, 6)$	AD	0.2522	1.2417
	$\alpha = 10605.33$	$\mathcal{H}_n(\sigma^*)$ with $\boldsymbol{\psi} = (0.99, 0.95)$	0.1767	0.0057
Pareto Tails	$\theta_{p_1} = 0.3738528$			
	$\theta_{p_2} = 4.0654859$			

The scale parameter σ has been estimated as $\sigma^* = 1.943354 \times 10^{-9}$ using (3.3.11), based on the fitted ME-P-P null distribution. As for the initial ME-P null, we again apply the KS, CvM, AD and $\mathcal{H}_n(\sigma)$ tests and the corresponding results are summarized in Table 3.5. As can be seen, at a significance level $p = 0.05$, the null is not rejected by all statistics, indicating a good fit. This is also confirmed by analyzing the PP and QQ plots, shown in Figure 3.13.

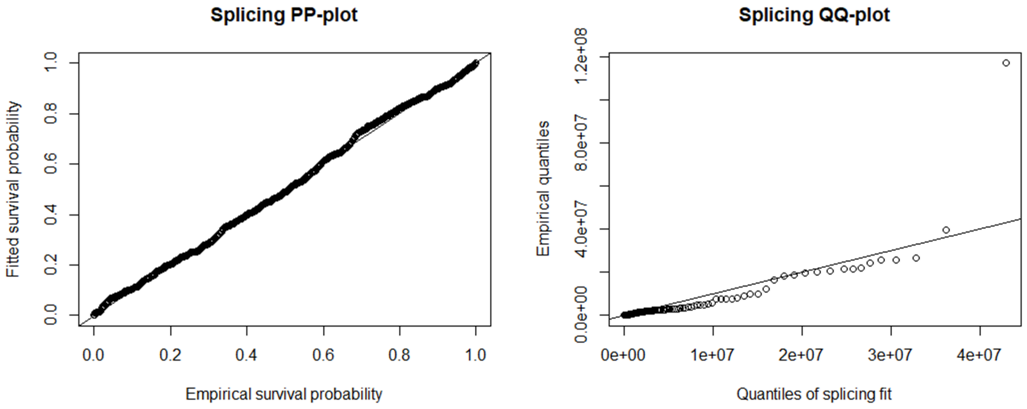


Figure 3.13: The PP and QQ plots for the ME-P-P distribution in Table 3.5

Finally, we slightly modified the ME-P-P null by changing C_1 to be the 90% quantile of the data in Group A. In addition to the KS, CvM, AD tests, we then considered two versions of the $\mathcal{H}_n(\sigma)$ test, $\mathcal{H}_n(\sigma_1^*)$, tuned for the tail with $\sigma_1^* = 1.869573 \times 10^{-9}$ and $\mathcal{H}_n(\sigma_2^*)$, tuned for the body with $\sigma_2^* = 5.382781 \times 10^{-6}$. Both σ_1^* and σ_2^* have been obtained following (3.3.11) with $\boldsymbol{\psi_1} = (0.99, 0.95)$ and $\boldsymbol{\psi_2} =$

$(0.6, 0.4)$ selected from Table 3.1 accordingly. We have repeatede the same estimation and test steps as before, and the results are summarized in Table 3.6 and illustrated in Figure 3.14.

Table 3.6: Parameters of the C_1 updated ME-P-P null fitted to Group A data for C_1 and C_2 equal to the 90% and 99% data quantiles and the results of the goodness-of-fit testing based on Group B data

Estimated Parameters of the ME-P-P null for Group A		Goodness of fit for Group B		
Splicing Point	$C_1 = 1,404,102$, $C_2 = 25,142,117$	**Statistics**	***p*-values**	**Values**
	$(\phi_1, \phi_2, \phi_3) = (0.9, 0.0892, 0.0108)$	KS	$< 2 \times 10^{-16}$	0.2249
Mixed Erlang	$\beta = (0.8717319, 0.1282681)$	CvM	$< 2 \times 10^{-16}$	11.981
$(M = 2)$	$r = (1, 13)$	AD	7.38×10^{-7}	82.262
	$\alpha = 50624.60$	$\mathcal{H}_n(\sigma_1^*)$ with $\boldsymbol{\psi} = (0.99, 0.95)$	0.4876	0.0040
Pareto Tails	$\theta_{p_1} = 0.5236715$			
	$\theta_{p_2} = 4.0654859$	$\mathcal{H}_n(\sigma_2^*)$ with $\boldsymbol{\psi} = (0.6, 0.4)$	1.94×10^{-7}	0.0783

As can be seen from Figure 3.14, the updated ME-P-P null fits the body poorly (as seen from the PP plot), but fits the right tail quite well (as seen from the QQ plot). The results from the goodness-of-fit testing summarized in Table 3.6 show that, the null is rejected by the KS, CvM, AD and the body sensitive $\mathcal{H}_n(\sigma_2^*)$ but it is not rejected by the right-tail sensitive $\mathcal{H}_n(\sigma_1^*)$, which is coherent with Figure 3.14.

In conclusion, with this final example, we demonstrate that under one and the same null distribution, it is possible to tune σ for different testing purposes, i.e. targeting differences expected either in the body or in the tail.

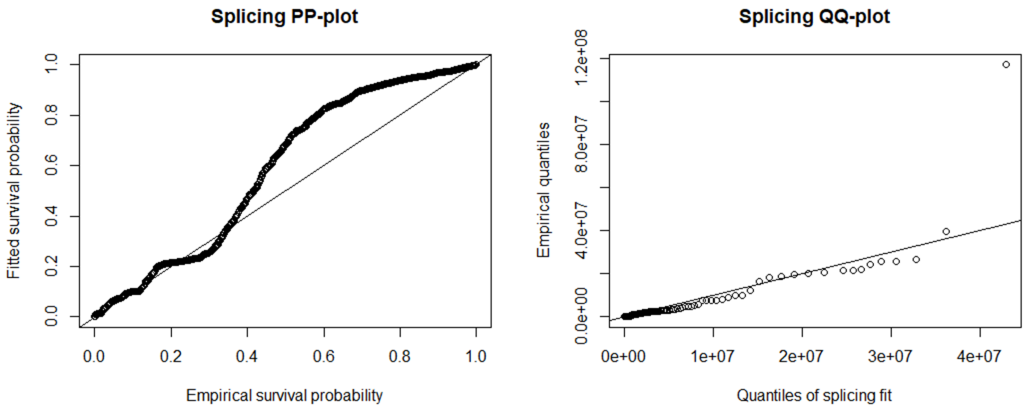


Figure 3.14: The PP and QQ plots for the C_1 updated ME-P-P distribution in Table 3.6

3.6 Discussion

We have proposed to use the Hausdorff metric to measure the distance between two (empirical) distribution functions and to view this metric as a goodness-of-fit test statistic, referred to as the H test. In Section 3.2, we have explored the properties of the H test and provided efficient algorithms to compute it and its exact/asymptotic p-value. The latter algorithms are very numerically efficient, which makes the H test particularly suitable for goodness-of-fit testing in data-rich environments such as machine learning, where large data samples are often collected.

We have shown in Section 3.3, that the H test can be tuned to become tail sensitive if certain convexity/concavity conditions with respect to the underlying null distribution are met, and we have also provided the corresponding tuning rules. These conditions are commonly met in data-scarce applications such as extreme value theory applied to finance and insurance, environmental science, cybersecurity, operations research, and many others. In such applications, observations are rare, as each of them is a large loss occurring with small (tail) probability, and the null distribution is often taken to be one of the common heavy-tailed distributions, or a related spliced construction. As illustrated in Section 3.5, the required convexity/-concavity conditions are met for such choices of the null distribution.

As illustrated in Sections 3.4 and 3.5, the power of the H test is improved for different tail alternatives, and the sample size required to reject the null i.e. to validate the model is smaller, as evidenced by its Bahadur efficiency, compared to other classical tests. These properties make the H test a more appropriate goodness-of-fit test alternative in such extreme value applications.

Appendix for Chapter 3

3.A Proofs for Chapter 3

Before we present some important results, it would be useful to give some auxiliary results and properties first.

As stated by (3.2.14), the Hausdorff metric H and the Lévy metric ρ_{Levy} coincide, which is also evidenced by their coincidental geometric interpretation as the side of the largest square inscribed between the corresponding planar curves. This fact has also been highlighted in Theorem 1* of Rachev (1984). Therefore, the properties of the Lévy metric are also valid for the metric H, as summarized by the Lemma 3.46, which is a restatement of properties (2) (6) and (7), summarized in Zolotarev (2011) with respect to the Lévy metric.

Lemma 3.46. *Suppose, F and G are nondecreasing, right-continuous functions. Denote by $F^{-1}(x) = \inf\{t : F(t) \geq x\}$ and $G^{-1}(x) = \inf\{t : G(t) \geq x\}$, the inverse functions of $F(x)$ and $G(x)$, and denote by F^{-1c} and G^{-1c}, their corresponding planar curves. We have*

$$H(F^c, G^c) = H(F^{-1c}, G^{-1c}), \tag{3.A.1a}$$

$$H^2(F^c, G^c) \leq \int_{-\infty}^{+\infty} |F(x) - G(x)| dx, \tag{3.A.1b}$$

$$H(F^c, G^c) \leq \sup_{x \in \mathbb{R}} |F(x) - G(x)| \leq H(F^c, G^c) + h(H(F^c, G^c)) \tag{3.A.1c}$$

where $h(x) = \min\{\sup_t |F(x+t) - F(t)|, \sup_t |G(x+t) - G(t)|\}$.

Let us note that the Hausdorff metric is defined with respect to two compact subset. Therefore, for the one-sample goodness-of-fit test, in the following, we give the localization consideration, which will be used in later proofs.

Lemma 3.47. *For $M > 0$, denote set $\mathcal{A}_M = [-M, M] \times [0, 1]$. Then*

$$\begin{aligned} H(F^c, F_n^c) &= \lim_{M\to\infty} H(F^c \cap \mathcal{A}_M, F_n^c \cap \mathcal{A}_M), \\ H(gr(F), gr(F_n)) &= \lim_{M\to\infty} H(gr(F) \cap \mathcal{A}_M, gr(F_n) \cap \mathcal{A}_M). \end{aligned} \tag{3.A.2}$$

Proof. Since $\lim_{x\to-\infty} F(x) = 0$, $\lim_{x\to+\infty} F(x) = 1$, $\forall\, \varepsilon > 0$, there exists $M > 0$, such that

$$F(x) < \varepsilon \text{ for any } x < -M \text{ and } F(x) > 1 - \varepsilon \text{ for any } x > M,$$
$$F_n(x) = 0 \text{ for any } x < -M \text{ and } F_n(x) = 1 \text{ for any } x > M$$

Let us define the function $\hat{F}$ as

$$\hat{F}(x) = \begin{cases} F(x) & |x| \le M \\ 0 & x < -M \\ 1 & x > M \end{cases}.$$

Then, $\forall\, x \in [-M, M]$, $(x, F(x)) \in F^c \cap \hat{F}^c$, which implies,

$$\inf_{B\in F^c} \rho_\infty((x, \hat{F}(x)), B) = 0 \text{ and } \inf_{A\in \hat{F}^c} \rho_\infty((x, F(x)), A) = 0.$$

$\forall\, |x| > M$, it is also easy to verify that

$$\inf_{B\in F^c} \rho_\infty((x, \hat{F}(x)), B) < \varepsilon \text{ and } \inf_{A\in \hat{F}^c} \rho_\infty((x, F(x)), A) < \varepsilon$$

Hence, $H(F^c, \hat{F}^c) < \varepsilon$. In addition, since $\hat{F}(x) = F_n(x)$ for $|x| > M$, we have

$$\begin{aligned} \sup_{A\in F^c\cap\mathcal{A}_M} \inf_{B\in F_n^c\cap\mathcal{A}_M} \rho_\infty(A, B) &= \sup_{A\in \hat{F}^c} \inf_{B\in F_n^c} \rho_\infty(A, B) \\ \sup_{A\in gr(F)\cap\mathcal{A}_M} \inf_{B\in gr(F_n)\cap\mathcal{A}_M} \rho_\infty(A, B) &= \sup_{A\in \hat{F}^c} \inf_{B\in gr(F_n)} \rho_\infty(A, B) \end{aligned}$$

Similarly, we can also show that

$$\begin{aligned} \sup_{B\in F_n^c\cap\mathcal{A}_M} \inf_{A\in F^c\cap\mathcal{A}_M} \rho_\infty(A, B) &= \sup_{B\in F_n^c} \inf_{A\in \hat{F}^c} \rho_\infty(A, B) \\ \sup_{B\in gr(F_n)\cap\mathcal{A}_M} \inf_{A\in gr(F)\cap\mathcal{A}_M} \rho_\infty(A, B) &= \sup_{B\in gr(F_n)} \inf_{A\in \hat{F}^c} \rho_\infty(A, B) \end{aligned}$$

Hence, we can show that

$$H(F^c \cap \mathcal{A}_M, F_n^c \cap \mathcal{A}_M) = H(\hat{F}^c, F_n^c)$$
$$H(gr(F) \cap \mathcal{A}_M, gr(F_n) \cap \mathcal{A}_M) = H(\hat{F}^c, gr(F_n))$$

Therefore, the triangle inequality implies that

$$\begin{aligned} &|H(F^c, F_n^c) - H(F^c \cap \mathcal{A}_M, F_n^c \cap \mathcal{A}_M)| \\ \leq &|H(F^c, \hat{F}^c) + H(\hat{F}^c, F_n^c) - H(F^c \cap \mathcal{A}_M, F_n^c \cap \mathcal{A}_M)| < \varepsilon \\ &|H(F^c, gr(F_n)) - H(F^c \cap \mathcal{A}_M, gr(F_n) \cap \mathcal{A}_M)| \leq H(F^c, \hat{F}^c) < \varepsilon \end{aligned}$$

Thus, the lemma follows due to the arbitrariness of ε. □

Lemma 3.47 can be extended to the planar curves of any distribution functions, but not to two arbitrary closed sets. Note that in a more general case, i.e. when the sets $\mathcal{A}$ and $\mathcal{B}$ are closed, and for an arbitrary sequence of compact sets $\mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3, \ldots$ such that $\mathcal{A}_l \to \mathcal{A}$ as $l = 1, 2, 3, \ldots$ with respect to the hit-or-miss topology, $H(\mathcal{A}_l, \mathcal{B})$ does not necessarily converge to $H(\mathcal{A}, \mathcal{B})$. Formally, the hit-or-miss topology is defined through the convergence criterion:

$$\mathcal{A}_l \to \mathcal{A} \text{ if and only if } \mathcal{A} = \cup_l \cap_i \mathcal{A}_{l+i} = \cap_l \cup_i \mathcal{A}_{l+i}.$$

This topology is defined on the collection of all closed subsets. The latter topology restricted on the collection of all the compact subsets is coarser than the myope topology, which is induced by the Hausdorff metric (cf. Matheron, 1974).

Furthermore, for arbitrary functions F and G, if $M_1 > M_2 > 0$, it is not necessarily true that

$$\begin{aligned} H\big(F^c \cap ([-M_1, M_1] \times \mathbb{R}), G^c \cap ([-M_1, M_1] \times \mathbb{R})\big) > \\ H\big(F^c \cap ([-M_2, M_2] \times \mathbb{R}), G^c \cap ([-M_2, M_2] \times \mathbb{R})\big). \end{aligned}$$

See Section 2.3 in Sendov and Beer (2012) for a counter-example.

Lemma 3.48. *For any $A \in \mathbb{R}^2$ and any distribution function F, $\inf_{B \in F^c} \rho_\infty(A, B)$ is always attainable. Furthermore, if $\inf_{B \in F^c} \rho_\infty(A, B) > 0$, let B_0 be the crossing point of F^c with the line passing through A and crossing the axis Ox at an angle*

$3\pi/4$, *then* $\rho_\infty(A, B_0) = \inf_{B\in F^c} \rho_\infty(A, B)$.

Proof. It is easy to prove that $\inf_{B\in F^c} \rho_\infty(A, B)$ is attainable, since the set $F^c \cap S(A, \inf_{B\in F^c} \rho_\infty(A, B) + \varepsilon)$ is nonempty and compact for arbitrary ε, where

$$S(A, \inf_{B\in F^c} \rho_\infty(A, B) + \varepsilon)$$

is the square with side $\inf_{B\in F^c} \rho_\infty(A, B) + \varepsilon$ and center at the point A.

The proof of $\rho_\infty(A, B_0) = \inf_{B\in F^c} \rho_\infty(A, B)$ is trivial, thus will be omitted. □

Proof of Lemma 3.5. Given the planar curve F_n, there exists $M_0 > 0$, such that $F_n^c/\mathcal{A}_{M_0} = gr(F_n)/\mathcal{A}_{M_0}$. Note that the set $F_n^c \cap \mathcal{A}_{M_0}$ is compact, and that according to Lemma 3.48, $\exists\, A_0 = (x_0, y_0) \in F_n^c \cap \mathcal{A}_{M_0}$ and $B_0 = (x_{B_0}, y_{B_0}) \in F^c$ such that,

$$H(F_n^c \cap \mathcal{A}_{M_0}, F^c \cap \mathcal{A}_M) = \inf_{B\in F^c} \rho_\infty(A_0, B) = \rho_\infty(A_0, B_0). \tag{3.A.3}$$

Furthermore, a line passing point A_0 and B_0 should cross the axis Ox at an angle of $\frac{3}{4}\pi$.

Hence, according to Lemma 3.47, to prove Lemma 3.5, it suffices to show that

$$H(F_n^c \cap \mathcal{A}_{M_0}, F^c \cap \mathcal{A}_M) = H(gr(F_n) \cap \mathcal{A}_{M_0}, gr(F) \cap \mathcal{A}_{M_0}). \tag{3.A.4}$$

In what follows, we will prove (3.A.4) by contradiction.

If (3.A.4) does not hold, then one must have that,

$$\forall\, A \in \overline{gr(F_n)} \cap \mathcal{A}_{M_0},\ \inf_{B\in F^c} \rho_\infty(A, B) < H(F_n^c \cap \mathcal{A}_{M_0}, F^c \cap \mathcal{A}_M). \tag{3.A.5}$$

Thus, there should exist $A_0 = (x_0, y_0) \in F_n^c/\overline{gr(F_n)}$, i.e. $F_n(x_0-) < y_0 < F_n(x_0)$, where $\overline{gr(F_n)}$ is the closure of $gr(F_n)$. In other words, the point A_0 should fall within one of the internal jump segments of F_n^c. Denote by $A_1 = (x_0, F_n(x_0-))$, $A_2 = (x_0, F_n(x_0))$. According to Lemma 3.48, there exist $B_1 = (x_{B_1}, y_{B_1}), B_2 = (x_{B_2}, y_{B_2}) \in F^c$ such that

$$\inf_{B\in F^c} \rho_\infty(A_1, B) = \rho_\infty(A_1, B_1) \text{ and } \inf_{B\in F^c} \rho_\infty(A_2, B) = \rho_\infty(A_2, B_2),$$

and that both the lines passing through the points, A_1, B_1 and A_2, B_2 cross the axis

Ox at an angle $\frac{3}{4}\pi$. Therefore, we have $|x_{B_1} - x_0| = \rho_\infty(A_1, B_1)$ and $|x_{B_2} - x_0| = \rho_\infty(A_2, B_2)$.

Since F is monotonic one should have $x_{B_1} \leq x_{B_0} \leq x_{B_2}$. Hence at least one of the $|x_{B_1} - x_0| = \rho_\infty(A_1, B_1)$ and $|x_{B_2} - x_0| = \rho_\infty(A_2, B_2)$ is not less than $|x_B - x_0| = \rho_\infty(A_0, B_0)$. This implies

$$\sup_{A \in \overline{gr(F_n)} \cap \mathcal{A}_{M_0}} \inf_{B \in F^c} \rho_\infty(A, B) \geq \rho_\infty(A_0, B_0) = H(F_n^c \cap \mathcal{A}_{M_0}, F^c \cap \mathcal{A}_M),$$

which contradicts with (3.A.5). Therefore, (3.A.4) holds. Similarly, we can also show that

$$\sup_{A \in F^c} \underline{\inf_{B \in gr(F) \cap \mathcal{A}_{M_0}}} \rho_\infty(A, B) = H(F_n^c \cap \mathcal{A}_{M_0}, F^c \cap \mathcal{A}_M).$$

Thus, the lemma holds. □

Proof of Lemma 3.6. Following Definition 3.3, we have

$$H(F^c, F_n^c) = \max\left[\sup_{A \in F_n^c} \inf_{B \in F^c} \rho_\infty(A, B), \sup_{B \in F^c} \inf_{A \in F_n^c} \rho_\infty(A, B)\right].$$

From Lemma 3.47, it is clear that sets F^c and F_n^c are locally compact. Thus, there exists $A^* \in F^c$, and $B^* \in F_n^c$, such that

$$\rho_\infty(A^*, B^*) = \inf_{B \in F_n^c} \rho_\infty(A^*, B) = \sup_{A \in F^c} \inf_{B \in F_n^c} \rho_\infty(A^*, B).$$

Furthermore, according to Lemma 3.48, the line passing through the points A^*, B^* crosses the axis Ox at an angle $3\pi/4$. Therefore $\rho_\infty(A^*, B^*) = \inf_{A \in F^c} \rho_\infty(A, B^*)$. Thus,

$$\sup_{A \in F^c} \inf_{B \in F_n^c} \rho_\infty(A, B) = \rho_\infty(A^*, B^*) \leq \sup_{B \in F_n^c} \inf_{A \in F^c} \rho_\infty(A, B).$$

By symmetry, we can also show that $\sup_{B \in F_n^c} \inf_{A \in F^c} \rho_\infty(A, B^*) \leq \sup_{A \in F^c} \inf_{B \in F_n^c} \rho_\infty(A, B)$. This completes the proof of the lemma. □

Lemma 3.6 allows us to consider only one of the two sup-inf forms in Definition 3.3 for the distance $H(F^c, F_n^c)$. Therefore it is very useful in deriving all the proofs and developing the numerical algorithms for evaluating the H test and its distributions.

Proof of Lemma 3.8. Given the sets $\mathcal{A}$ and $\mathcal{B}$, denote by $\mathcal{A}\oplus\mathcal{B}$ their Minkowski sum, i.e. $\mathcal{A}\oplus\mathcal{B}=\{P_1+P_2 : P_1\in\mathcal{A}, P_2\in\mathcal{B}\}$. Denote by $O=(0,0)$. Then an alternative expression of the Hausdorff distance is (cf. Sendov and Beer, 2012)

$$H(F^c,F_n^c)=\inf\{\epsilon : F^c\subset S(O,2\epsilon)\oplus F_n^c, F_n^c\subset S(O,2\epsilon)\oplus F^c\}. \tag{3.A.6}$$

For simplicity, denote $d_0=H(F^c,F_n^c)<\infty$. Then $\forall\,\varepsilon>0$, we have

$$F^c\subset S(O,2(d_0+\varepsilon))\oplus F_n^c,\ F_n^c\subset S(O,2(d_0+\varepsilon))\oplus F^c. \tag{3.A.7}$$

Note that both $\mathbb{R}^2/(S(O,2(d_0+\varepsilon))\oplus F_n^c)$ and $\mathbb{R}^2/(S(O,2(d_0+\varepsilon))\oplus F^c)$ are the unions of two disjoint but connected open sets. These open sets can be viewed as an epigraph and a hypograph of a certain (interval-valued) function. For simplicity, denote by $\mathcal{U}_{F^c}(\varepsilon)$ and $\mathcal{U}_{F_n^c}(\varepsilon)$ the epigraphs of the sets/functions $S(O,2\varepsilon)\oplus F^c$ and $S(O,2\varepsilon)\oplus F_n^c$ respectively, i.e.

$$\begin{aligned}\mathcal{U}_{F^c}(\varepsilon)&=\{(x,y):\forall\,(x,z)\in S(O,2\varepsilon)\oplus F^c,\ y>z\},\\ \mathcal{U}_{F_n^c}(\varepsilon)&=\{(x,y):\forall\,(x,z)\in S(O,2\varepsilon)\oplus F_n^c,\ y>z\}.\end{aligned}$$

Denote by $\mathcal{L}_{F^c}(\varepsilon)$ and $\mathcal{L}_{F_n^c}(\varepsilon)$ their corresponding hypographs. Clearly

$$\begin{aligned}\mathbb{R}^2/(S(O,2\varepsilon)\oplus F^c)&=\mathcal{U}_{F^c}(\varepsilon)\cup\mathcal{L}_{F^c}(\varepsilon),\\ \mathbb{R}^2/(S(O,2\varepsilon)\oplus F_n^c)&=\mathcal{U}_{F_n^c}(\varepsilon)\cup\mathcal{L}_{F_n^c}(\varepsilon).\end{aligned}$$

Furthermore, from (3.A.7), we have

$$\mathcal{U}_{F^c}(d_0+\varepsilon)\subset\mathcal{U}_{F^c},\ \mathcal{L}_{F^c}(d_0+\varepsilon)\subset\mathcal{L}_{F_n^c},\ \mathcal{U}_{F_n^c}(d_0+\varepsilon)\subset\mathcal{U}_{F_n^c},\ \mathcal{L}_{F_n^c}(d_0+\varepsilon)\subset\mathcal{L}_{F_n^c}.$$

Hence

$$\begin{aligned}&\Big[(\mathcal{U}_{F_n^c}(d_0+\varepsilon)\cap\mathcal{U}_{F^c}(d_0+\varepsilon))\cup(\mathcal{L}_{F_n^c}(d_0+\varepsilon)\cap\mathcal{L}_{F^c}(d_0+\varepsilon))\Big]\\ &\qquad\subset\left[(\mathcal{U}_{F_n^c}\cap\mathcal{U}_{F^c})\cup(\mathcal{L}_{F_n^c}\cap\mathcal{L}_{F^c})\right].\end{aligned} \tag{3.A.8}$$

Therefore, with respect to the complementary sets of both sides of (3.A.8), we have

$$\mathbb{R}^2/\Big\{(\mathcal{U}_{F_n^c}\cap\mathcal{U}_{F^c})\cup(\mathcal{L}_{F_n^c}\cap\mathcal{L}_{F^c})\Big\}=\mathcal{G}\subset$$
$$\mathbb{R}^2/\Big\{(\mathcal{U}_{F_n^c}(d_0+\varepsilon)\cap\mathcal{U}_{F^c}(d_0+\varepsilon))\cup(\mathcal{L}_{F_n^c}(d_0+\varepsilon)\cap\mathcal{L}_{F^c}(d_0+\varepsilon))\Big\}$$
$$=(S(O,2(d_0+\varepsilon))\oplus F_n^c)\cap(S(O,2(d_0+\varepsilon))\oplus F_n^c).$$

Thus, if $S(P,d)\subset\mathcal{G}$ then $d<d_0+\varepsilon$. Inequality (3.2.9) follows by the arbitrariness of ε.

Then according to Lemmas 3.47 and 3.48, there exist points $A_1\in F^c$ and $B_1\in F_n^c$ such that

$$\rho_\infty(A_1,B_1)=\inf_{A\in F^c}\rho_\infty(A,B_1)=d_0,$$

and the line passing through the points A_1 and B_1 crosses the axis Ox at an angle $3\pi/4$. Clearly, $B_1\in S(A_1,2d_0)$ and $A_1\in S(B_1,2d_0)$. From (3.A.6), for any $\epsilon>0$, it is easy to show that $S(A_1,2d_0-2\epsilon)\cap S(B_1,2d_0-2\epsilon)=S(\frac{A_1+B_1}{2},d_0-\epsilon)\subset\mathcal{G}$. Hence, we have $d_0-\epsilon\le\sup\{d:S(P,d)\subset\mathcal{G}\}\le d_0$. Equality (3.2.10) follows by the arbitrariness of ϵ. □

Proof of Lemma 3.9. To show that the supremum can be taken only with respect to the vertices of F_n^c, following (3.A.4), it suffices to show that the supremum of

$$\inf_{B\in F^c}\rho_\infty(A,B)$$

with respect to $A\in F_n^c$ cannot be achieved on the interior of the horizontal segments between the vertices A_{2i-1} and A_{2i}, $i=1,2,\ldots,n$. This proof is similar to proving (3.A.4) in Lemma 3.5, so will be omitted. Therefore we have

$$H(F^c,F_n^c)=\max_{B\in\{A_i\}_{i=1}^{2n}}\inf_{A\in F^c}\rho_\infty(B,A). \tag{3.A.9}$$

Since the maximum in (3.A.9) cannot be achieved at the vertices that are locally closest to F^c (cf. Fig. 3.3), (3.A.9) can be rewritten as

$$H(F^c,F_n^c)=\max_{B\in\mathcal{B}_{loc}}\inf_{A\in F^c}\rho_\infty(B,A).$$

By applying Lemma 3.48, we have that for every $B_l \in \mathcal{B}_{loc}$, $\inf_{A\in F^c} \rho_\infty(B_l, A) = \rho_\infty(B_l, E_l)$. Therefore, (3.2.15) follows. □

Proof of Theorem 3.11. Consider vertical straight lines $\mathcal{L}'_l$ perpendicular to the axis Ox, that pass through each of the vertices B_l, $l = 1,2,\ldots,\nu$. Denote by Q_l, $l = 1,2,\ldots,\nu$ the points of intersection of the lines $\mathcal{L}'_l$ with the planar curve F^c. Then it is easy to verify that

$$\rho_D(F, F_n) = \max\{\rho_\infty(B_l, Q_l), l = 1,2,\ldots,\nu\}. \tag{3.A.10}$$

Denote by x_{Q_l} and y_{Q_l}, the x and y coordinates of the points Q_i, $l = 1,\ldots,\nu$.

If $B_l \in \mathcal{U}_{F^c}$, then we have $x_{B_l} = x_{Q_l} < x_{E_l}$, Therefore, since $F(x)$ is nondecreasing, $y_{Q_l} \le y_{E_l} < y_{B_l}$ and $\rho_\infty(B_l, Q_l) = y_{B_l} - y_{Q_l} \ge y_{B_l} - y_{E_l} = \rho_\infty(B_l, E_l)$.

If $B_l \in \mathcal{L}_{F^c}$, $x_{E_l} < x_{B_l} = x_{Q_l}$. For the same reason, $y_{B_l} < y_{E_l} \le y_{Q_l}$ and $\rho_\infty(B_l, Q_l) = y_{Q_l} - y_{B_l} \ge y_{E_l} - y_{B_l} = \rho_\infty(B_l, E_l)$.

In summary, $\rho_\infty(B_l, Q_l) \ge \rho_\infty(B_l, E_l)$, $l = 1,2,\ldots,\nu$ and result (3.2.17) follows from (3.2.15) and (3.A.10). □

Remark 3.49. It should be pointed out that, the result of Theorem 3.11 coincides with the first inequality in (3.A.1c) (see Lemma 3.46).

Proof of Lemma 3.13. The proof of Lemma 3.13 can be directly obtained by Lemma 3.48, thus will be omitted. □

Proof of Lemma 3.14. (Proof by contradiction) Suppose $\max_\lambda d_\lambda < H(F^c, F_n^c)$. It implies that there exist a point $P_2 = (x'_2, y'_2) \in \mathbb{R}^2$ and a constant $d_2 \in \mathbb{R}$, such that $\max_\lambda d_\lambda < d_2 < H(F^c, F_n^c)$ and $S(P_2, d_2) \subset \mathcal{G}$.

Let $\lambda_0 = x'_2 + y'_2$, it is not difficult to verify that the point $P_{\lambda_0} = (x_{\lambda_0}, y_{\lambda_0})$ which is the center of the square S_{λ_0}, P_2, $(x'_2 - d_2/2, y'_2 + d_2/2)$ and the point $(x_{\lambda_0} - d_{\lambda_0}/2, y_{\lambda_0} + d_{\lambda_0}/2)$ are on the line $x + y = \lambda_0$. Furthermore, we have that $(x_{\lambda_0} - d_{\lambda_0}/2, y_{\lambda_0} + d_{\lambda_0}/2) \in F^c \cup F_n^c$. And since $S(P_2, d_2) \subset \mathcal{G}$, one must have $(x'_2 - d_2/2, y'_2 + d_2/2) \in S(P_{\lambda_0}, d_{\lambda_0})$. By applying the same logic, $(x'_2 + d_2/2, y'_2 - d_2/2) \in R(P_{\lambda_0}, d_{\lambda_0})$. Hence, $d_2 \le d_{\lambda_0}$. However, this contradicts with $\sup_\lambda d_\lambda < d_2$, which completes the proof of Lemma 3.14. □

The statement of Lemma 3.14 is equivalent to (2.2) in Rachev (1981), formulated

with respect to the Lévy metric. Lemma 3.14 is a further generalization of Lemma 3.9, which is needed to express the Hausdorff metric as an appropriate supremum under the transformed coordinate system, see Section 3.2.3. Before proving Theorem 3.17, it will be useful to recall Example 3.16 (i), (iii) and the corresponding graphical illustrations of the image sets of $S(P,d)$ and F_n^c given in Figure 3.5.

Remark 3.50. Lemmas 3.13 and 3.14, are also true for arbitrary right continuous nondecreasing functions F' and G', as substitutes of F and F_n, respectively.

Proof of Theorem 3.17. For any real number λ, consider the image set of S_λ in the uOv plane, it is easy to see that $T(S_\lambda)$ is a parallelogram with both its base and height equal to λ (cf. Example 3.16 and Figure 3.5). According to the definition of S_λ, $T(S_\lambda)\cap T(F^c)$ and $T(S_\lambda)\cap T(F_n^c)$ are non empty. In addition, denote by O_λ and R_λ correspondingly the upper-left and lower-right vertices of the square S_λ. Clearly, $O_\lambda, R_\lambda \in F^c \cup F_n^c$. Then it is not difficult to observe that

$$\rho_\infty(O_\lambda, R_\lambda) = d_\lambda = |K(\lambda) - K_n(\lambda)|. \tag{3.A.11}$$

Equality (3.2.20) now follows by taking the supremum on both sides of the last equality in (3.A.11) and applying Lemma 3.14. The fact that the supremum in (3.2.20) is achieved either at the point $t = X_{(i)} + \frac{i-1}{n}$ or at $t = X_{(i)} + \frac{i}{n}$ follows from (3.A.9) (see the proof of Lemma 3.9). This completes the proof of Theorem 3.17. □

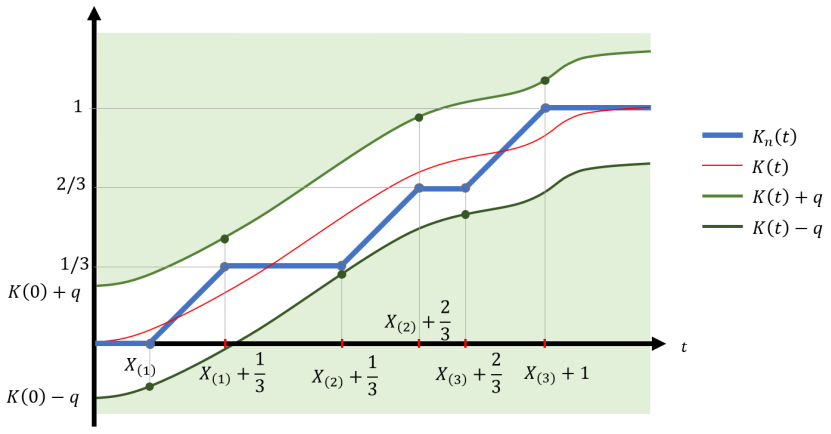


Figure 3.15: Illustration of the equivalence of $\mathbb{P}(K(t)-q \le K_n(t) \le K(t)+q, \text{for all } t)$ to $\mathbb{P}\left(K^{-1}(\frac{i}{n}-q)-\frac{i}{n} \le X_{(i)} \le K^{-1}(\frac{i-1}{n}+q)-\frac{i-1}{n}, \text{for } 1\le i\le n\right)$ (cf. Proof of Theorem 3.19), for $n=3$.

Proof of Theorem 3.19. We have

$$
\begin{aligned}
&\mathbb{P}(H(F^c,F_n^c)>q)=\mathbb{P}(\sup_{-\infty<t<+\infty}|K_n(t)-K(t)|>q)\\
=&1-\mathbb{P}(\sup_{-\infty<t<+\infty}|K_n(t)-K(t)|\le q)\\
=&1-\mathbb{P}(|K_n(t)-K(t)|\le q,\text{for all } t)\\
=&1-\mathbb{P}(K(t)-q\le K_n(t)\le K(t)+q,\text{for all } t)\\
=&1-\mathbb{P}\bigg(K(X_{(i)}+\frac{i-1}{n})-q\le K_n(X_{(i)}+\frac{i-1}{n}) \text{ and}\\
&K_n(X_{(i)}+\frac{i}{n})\le K(X_{(i)}+\frac{i}{n})+q,\text{for } 1\le i\le n\bigg)\\
=&1-\mathbb{P}\bigg(K(X_{(i)}+\frac{i-1}{n})\le\frac{i-1}{n}+q\\
&\text{and } \frac{i}{n}-q\le K(X_{(i)}+\frac{i}{n}),\text{for } 1\le i\le n\bigg)\\
=&1-\mathbb{P}\bigg(K^{-1}(\frac{i}{n}-q)-\frac{i}{n}\le X_{(i)}\le K^{-1}(\frac{i-1}{n}+q)-\frac{i-1}{n},\text{for } 1\le i\le n\bigg)\\
=&1-\mathbb{P}\bigg(F\Big(K^{-1}(\frac{i}{n}-q)-\frac{i}{n}\Big)\le U_{(i)}\le F\Big(K^{-1}(\frac{i-1}{n}+q)-\frac{i-1}{n}\Big),\text{for } 1\le i\le n\bigg)
\end{aligned}
$$

The equivalence of $\mathbb{P}(K(t)-q\le K_n(t)\le K(t)+q,\text{for all } t)$ to $\mathbb{P}\Big(K^{-1}(\frac{i}{n}-q)-\frac{i}{n}\le X_{(i)}\le K^{-1}(\frac{i-1}{n}+q)-\frac{i-1}{n},\text{for } 1\le i\le n\Big)$ is also graphically illustrated in Figure 3.15. Then, from the definition of $K(x)$, we have that $K^{-1}(x)=x+F^{-1}(x)$. Hence,

$$
\begin{aligned}
F\Big(K^{-1}(\frac{i}{n}-q)-\frac{i}{n}\Big)&=F(F^{-1}(\frac{i}{n}-q)-q),\\
F\Big(K^{-1}(\frac{i-1}{n}+q)-\frac{i}{n}\Big)&=F(F^{-1}(\frac{i-1}{n}+q)+q).
\end{aligned}
$$

□

Proof of Theorem 3.21. Following equality (5) in Dimitrova et al. (2020), one can re-express (3.2.21) as

$$
\mathbb{P}(\sqrt{n}H(F_n^c,F^c)\ge x)=1-\mathbb{P}\left(g_n(t)\le\eta_n(t)\le h_n(t),\forall\ 0\le t\le 1\right), \tag{3.A.12}
$$

where the process $\eta_n(t)=\sum_{i=1}^n\mathbb{1}\{U_i\le t\}$, $t\in[0,1]$ and the upper and lower boundary

functions $h_n(t)$, $g_n(t)$ are defined as

$$h_n(t) = \sum_{i=1}^{n} \mathbb{1}(a_{i,n} < t), \quad g_n(t) = \sum_{i=1}^{n} \mathbb{1}(b_{i,n} \le t), \tag{3.A.13}$$

and

$$\begin{aligned} a_{i,n} &= F(F^{-1}(\frac{i}{n} - xn^{-1/2}) - xn^{-1/2}), \\ b_{i,n} &= F(F^{-1}(\frac{i-1}{n} + xn^{-1/2}) + xn^{-1/2}), \end{aligned}$$

with $x = qn^{1/2}$. For simplicity, we denote the inverse function of $\frac{1}{\sqrt{n}}h_n(t)$ as $\alpha_n(u)$, $u \in [0, \sqrt{n}]$. Then from the definition of $h_n(t)$ in (3.A.13) for $\alpha_n(u)$, we have

$$\alpha_n(u) = a_{[\sqrt{n}u],n} = F(F^{-1}(\frac{[\sqrt{n}u]}{n} - xn^{-1/2}) - xn^{-1/2}), \tag{3.A.14}$$

where $[\cdot]$ denotes the integer part and $u \in [0, \sqrt{n}]$. Let $y = \frac{1}{\sqrt{n}}h_n(t) - \sqrt{n}t$. Since $\alpha_n(u)$ is the inverse function of $\frac{1}{\sqrt{n}}h_n(t)$, we have

$$t = \alpha_n(u), \; u = y + \sqrt{n}t,$$

and for large n, (3.A.14) can accordingly be re-written as

$$t = F(F^{-1}(n^{-1/2}(y-x) + t) - xn^{-1/2}),$$

or equivalently as

$$y = \sqrt{n}[F(F^{-1}(t) + xn^{-1/2}) - t] + x. \tag{3.A.15}$$

When $n \to \infty$, by the definition of the derivative, the right-hand side of (3.A.15) converges to $x(1 + f(F^{-1}(t)))$. Hence, for a fixed t, we have

$$\begin{aligned} \left(\frac{1}{\sqrt{n}}h_n(t) - \sqrt{n}t\right) &\to x\left(1 + f\left(F^{-1}(t)\right)\right), \\ \left(\frac{1}{\sqrt{n}}g_n(t) - \sqrt{n}t\right) &\to -x\left(1 + f(F^{-1}(t))\right). \end{aligned} \tag{3.A.16}$$

Since f is bounded, one can show that (3.A.16) are uniformly convergent.

For simplicity, denote by $X_n(t) = \sqrt{n}(\frac{1}{n}\eta_n(t) - t)$. Then by the Donsker's theorem, there exists a Brownian bridge $\mathbb{B}_0(t)$, $t \in [0,1]$ with $\mathbb{B}_0(0) = \mathbb{B}_0(0) = 0$,

$E[\mathbb{B}_0(t)] = 0$, $E[\mathbb{B}_0(t)\mathbb{B}_0(s)] = s(1-t)$ for $0 < s < t < 1$, such that

$$X_n(t) \xrightarrow{w} \mathbb{B}_0(t),$$

on the space $L^\infty[0,1]$, where $L^\infty[0,1]$ is the collection of functions defined on $[0,1]$ that are essentially bounded and $\xrightarrow{w}$ stands for weak convergence. We can then rewrite (3.A.12) as

$$\begin{aligned}&\mathbb{P}(\sqrt{n}H(F_n^c, F^c) \geq x)\\ =&1 - \mathbb{P}\left(\frac{1}{\sqrt{n}}g_n(t) - \sqrt{n}t \leq X_n(t) \leq \frac{1}{\sqrt{n}}h_n(t) - \sqrt{n}t, \forall\ 0 \leq t \leq 1\right).\end{aligned} \tag{3.A.17}$$

Denote by

$$\begin{aligned}\mathcal{A}_n &= \{z(t) \in L^\infty[0,1] : \frac{1}{\sqrt{n}}g_n(t) - \sqrt{n}t \leq z(t) \leq \frac{1}{\sqrt{n}}h_n(t) - \sqrt{n}t, \forall\ 0 \leq t \leq 1\},\\ \mathcal{A} &= \{z(t) \in L^\infty[0,1] : -x(1+f(F^{-1}(t))) \leq z(t) \leq x(1+f(F^{-1}(t))), \forall\ 0 \leq t \leq 1\},\end{aligned}$$

which represent subsets of $L^\infty[0,1]$. In view of (3.A.17), to estabilish (3.2.23), it suffices to show that

$$\lim_{n\to\infty} \mathbb{P}(X_n(t) \in \mathcal{A}_n) = \mathbb{P}(\mathbb{B}_0(t) \in \mathcal{A}). \tag{3.A.18}$$

In order to prove (3.A.18), for any $\varepsilon > 0$, let us define

$$\begin{aligned}\mathcal{A}^\varepsilon = \Big\{z(t) : &\inf_{t\in[0,1]}\Big[z(t) + x(1+f(F^{-1}(t)))\Big] \geq \varepsilon\\ &\text{and } \inf_{t\in[0,1]}\Big[x(1+f(F^{-1}(t))) - z(t)\Big] \geq \varepsilon\Big\},\\ \mathcal{A}^{-\varepsilon} = \Big\{z(t) : &\inf_{t\in[0,1]}\Big[z(t) + x(1+f(F^{-1}(t)))\Big] > -\varepsilon\\ &\text{and } \inf_{t\in[0,1]}\Big[x(1+f(F^{-1}(t))) - z(t)\Big] > -\varepsilon\Big\}.\end{aligned}$$

The set $\mathcal{A}^\varepsilon$ is closed in $L^\infty[0,1]$, whereas the set $\mathcal{A}^{-\varepsilon}$ is a Borel set with closure

$$\begin{aligned}\overline{\mathcal{A}^{-\varepsilon}} = \Big\{z(t) : &\inf_{t\in[0,1]}\Big[z(t) + x(1+f(F^{-1}(t)))\Big] \geq -\varepsilon\\ &\text{and } \inf_{t\in[0,1]}\Big[x(1+f(F^{-1}(t))) - z(t)\Big] \geq -\varepsilon\Big\}.\end{aligned}$$

Furthermore, due to the uniform convergence in (3.A.16), $\forall\, \varepsilon > 0$, there exists N such that,

$$\forall\, n > N,\ \mathcal{A}^{\varepsilon} \subset \mathcal{A}_n \subset \mathcal{A}^{-\varepsilon}. \tag{3.A.19}$$

Since $\mathcal{A}^{\varepsilon}$ is closed, for any $\varepsilon > 0$,

$$\begin{aligned}
\partial\mathcal{A}^{\varepsilon} \subset &\left\{z(t) : \inf_{t\in[0,1]}\Big[z(t) + x(1+f(F^{-1}(t)))\Big] = \varepsilon\right\} \cup \\
&\left\{\inf_{t\in[0,1]}\Big[x(1+f(F^{-1}(t))) - z(t)\Big] = \varepsilon\right\}, \\
\partial\mathcal{A}^{-\varepsilon} \subset &\left\{z(t) : \inf_{t\in[0,1]}\Big[z(t) + x(1+f(F^{-1}(t)))\Big] = -\varepsilon\right\} \cup \\
&\left\{\inf_{t\in[0,1]}\Big[x(1+f(F^{-1}(t))) - z(t)\Big] = -\varepsilon\right\},
\end{aligned}$$

where $\partial\mathcal{A}^{\varepsilon}$ and $\partial\mathcal{A}^{-\varepsilon}$ denote the boundary of $\mathcal{A}^{\varepsilon}$ and $\mathcal{A}^{-\varepsilon}$. By the property of the Brownian bridge,

$$\mathbb{P}(\mathbb{B}_0(t) \in \partial\mathcal{A}^{\varepsilon}) = \mathbb{P}(\mathbb{B}_0(t) \in \partial\mathcal{A}^{-\varepsilon}) = 0.$$

Therefore, by the Portmanteau theorem (cf. Theorem 1.3.4 in van der Vaart and Wellner, 2023), we have

$$\begin{aligned}
\lim_{n\to\infty} \mathbb{P}(X_n(t) \in \mathcal{A}^{\varepsilon}) = \mathbb{P}(\mathbb{B}_0(t) \in &\mathcal{A}^{\varepsilon}) \text{ and} \\
\lim_{n\to\infty} \mathbb{P}(X_n(t) \in \mathcal{A}^{-\varepsilon}) &= \mathbb{P}(\mathbb{B}_0(t) \in \mathcal{A}^{-\varepsilon})
\end{aligned} \tag{3.A.20}$$

holds for any $\varepsilon > 0$. By (3.A.19), $\forall\, \varepsilon > 0$, there exists N,

$$\forall\, n > N,\ \mathbb{P}(X_n(t) \in \mathcal{A}^{\varepsilon}) \le \mathbb{P}(X_n(t) \in \mathcal{A}_n) \le \mathbb{P}(X_n(t) \in \mathcal{A}^{-\varepsilon}).$$

We take the limit with respect to n and have

$$\begin{aligned}
\mathbb{P}(\mathbb{B}_0(t) \in \mathcal{A}^{\varepsilon}) &\le \liminf_{n\to\infty} \mathbb{P}(X_n(t) \in \mathcal{A}_n) \\
&\le \limsup_{n\to\infty} \mathbb{P}(X_n(t) \in \mathcal{A}_n) \le \mathbb{P}(\mathbb{B}_0(t) \in \mathcal{A}^{-\varepsilon}).
\end{aligned} \tag{3.A.21}$$

Note that when $\varepsilon \downarrow 0$, $\mathcal{A}^{\varepsilon} \uparrow \mathcal{A}^{\circ}$, and $\mathcal{A}^{-\varepsilon} \downarrow \mathcal{A}$, where $\mathcal{A}^{\circ}$ is the interior of $\mathcal{A}$, i.e.

$$\mathcal{A}^{\circ} = \{z(t) \in L^{\infty}[0,1] : -x(1+f(F^{-1}(t))) < z(t) < x(1+f(F^{-1}(t))), \forall\ 0 \le t \le 1\}.$$

Similarly, we can show that $\mathbb{P}(\mathbb{B}_0(t) \in \mathcal{A}^o) = \mathbb{P}(\mathbb{B}_0(t) \in \mathcal{A})$. Thus, by the arbitrariness of ε in (3.A.21),

$$\begin{aligned}\mathbb{P}(\mathbb{B}_0(t) \in \mathcal{A}) = \sup_{\varepsilon>0} \mathbb{P}(\mathbb{B}_0(t) \in \mathcal{A}^\varepsilon) &\leq \liminf_{n\to\infty} \mathbb{P}(X_n(t) \in \mathcal{A}_n) \\ &\leq \limsup_{n\to\infty} \mathbb{P}(X_n(t) \in \mathcal{A}_n) \leq \inf_{\varepsilon>0} \mathbb{P}(\mathbb{B}_0(t) \in \mathcal{A}^{-\varepsilon}) = \mathbb{P}(\mathbb{B}_0(t) \in \mathcal{A}),\end{aligned}$$

which yields (3.A.18). This completes the proof of (3.2.23) and therefore Theorem 3.21. □

Proof of Corollary 3.22. Since $F \sim Exp(\alpha)$, $f(F^{-1}(t)) = \alpha(1-t)$ for $0 \leq t \leq 1$. Note that these boundaries are linear, to obtain the result, we apply Theorem 4.2 of Anderson (1960), which states that

$$\mathbb{P}\left(-\gamma(t) \leq \mathbb{B}_0(t) \leq \gamma(t), \forall\ 0 \leq t \leq 1\right) = 1 - 2\sum_{k=1}^{\infty} (-1)^{k-1} e^{-2\delta(\delta+\beta)k^2} \tag{3.A.22}$$

where $\gamma(t) = \delta + \beta t$, noting that the bounds $\pm x(1+\alpha(1-t))$ in (3.2.24) are in the form of $\pm\gamma(t)$, with $\delta = x(1+\alpha)$, $\beta = -x\alpha$. □

Proof of Theorem 3.23. Let us recall the alternative expressions of $H(F_n^c(\omega), F^c)$ in (3.2.15) and $\mathcal{D}_n(\omega)$ in (3.A.10). Equalities (3.2.25) and (3.2.26) directly follow from the fact that $\rho_\infty(B_l, Q_l) = \frac{b}{1+b}\rho_\infty(B_l, E_l)$ for every $l = 1, \ldots, \nu$, when $F \sim U(a, a+b)$. Equality (3.2.27) follows from (3.A.22), replacing $\gamma(t) \equiv \frac{1+b}{b}x$. □

Proof of Proposition 3.26. Let us first note that the following two events are equivalent, i.e.

$$\{P_n(\sigma) > p\} = \{F_\sigma(x - q_\sigma^*) - q_\sigma^* \leq F_{n,\sigma}(x) \leq F_\sigma(x + q_\sigma^*) + q_\sigma^* \text{ for all } x\}, \tag{3.A.23}$$

where F_σ and $F_{n,\sigma}$ are defined in (3.3.1). The equivalence stems from $\{P_n(\sigma) > p\} = \{\mathcal{H}_n(\sigma) < q_\sigma^*\}$ and Theorem 3.17. Then to obtain (3.3.7), we need to apply the inverted scale $1/\sigma$ to the cdfs $F_{n,\sigma}(x,\omega)$, $F_\sigma(x - q_\sigma^*) - q_\sigma^*$ and $F_\sigma(x + q_\sigma^*) + q_\sigma^*$ in (3.A.23) so that they transform back to the original scale. Hence, $F_{n,\sigma}(x,\omega)$ becomes $F_n(x,\omega)$, and the latter boundaries then become $F_\sigma(\sigma x - q_\sigma^*) - q_\sigma^* = \tilde{F}_-(x;\sigma,p)$ and $F_\sigma(\sigma x + q_\sigma^*) + q_\sigma^* = \tilde{F}_+(x;\sigma,p)$. □

Proof of Proposition 3.27. Equation (3.3.8) follows from (3.3.3) and (3.3.4). □

Proof of Lemma 3.28. Let us prove the second inequality in (4.4.4), i.e. $\frac{q^*_{\sigma_2}}{\sigma_2} > \frac{q^*_{\sigma_1}}{\sigma_1}$ for $\sigma_1 > \sigma_2$. We show this holds by contradiction.

Suppose that $\frac{q^*_{\sigma_2}}{\sigma_2} \leq \frac{q^*_{\sigma_1}}{\sigma_1}$. According to Proposition 3.34, when $\mathcal{H}_n(\sigma_1) > q^*_{\sigma_1}$, there exists P_0, such that $S(P_0, q^*_{\sigma_1}; \sigma_1) \subset \mathcal{G}$. Since $\frac{q^*_{\sigma_2}}{\sigma_2} \leq \frac{q^*_{\sigma_1}}{\sigma_1}$, it is easy to see that $S(P_0, q^*_{\sigma_2}; \sigma_2) \subsetneq S(P_0, q^*_{\sigma_1}; \sigma_1) \subset \mathcal{G}$. The latter implies $\mathcal{H}_n(\sigma_2) > q^*_{\sigma_2}$. Hence, we have

$$\{\omega : \mathcal{H}_n(\sigma_1, \omega) > q^*_{\sigma_1}\} \subsetneq \{\omega : \mathcal{H}_n(\sigma_2, \omega) > q^*_{\sigma_2}\}. \tag{3.A.24}$$

Taking probabilites on both sides of (3.A.24), we have

$$\mathbb{P}\{\omega : \mathcal{H}_n(\sigma_1, \omega) > q^*_{\sigma_1}\} < \mathbb{P}\{\omega : \mathcal{H}_n(\sigma_2, \omega) > q^*_{\sigma_2}\}. \tag{3.A.25}$$

Noting the definition of $q^*(\sigma)$ in (3.3.6), we have $\mathbb{P}(P_n(\sigma_i) > p) = \mathbb{P}\{\omega : \mathcal{H}_n(\sigma_i, \omega) > q^*_{\sigma_i}\}$ under the null for $i = 1, 2$. Therefore (3.A.25) can be rewritten as

$$\mathbb{P}(P_n(\sigma_1) > p) < \mathbb{P}(P_n(\sigma_2) > p). \tag{3.A.26}$$

However, under the null hypothesis, (3.3.5) must hold, which contradicts with (3.A.26). One arrives at a similar contradiction in the case when $q^*_{\sigma_1} \leq q^*_{\sigma_2}$. This completes the proof of Lemma 3.28. □

Proof of Theorem 3.29. $\iota(x, \sigma) = F(x + \frac{q^*_\sigma}{\sigma}) - F(x - \frac{q^*_\sigma}{\sigma}) + 2q^*_\sigma$. For $\sigma_1 > \sigma_2$, inequality (4.4.4) suggests $q^*_{\sigma_1} - q^*_{\sigma_2} > 0$, $\frac{q^*_{\sigma_2}}{\sigma_2} > \frac{q^*_{\sigma_1}}{\sigma_1}$. Thus,

$$\begin{aligned} \iota(x, \sigma_1) - \iota(x, \sigma_2) =& F(x + \frac{q^*_{\sigma_1}}{\sigma_1}) - F(x + \frac{q^*_{\sigma_2}}{\sigma_2}) + \\ & \underbrace{F(x - \frac{q^*_{\sigma_2}}{\sigma_2}) - F(x - \frac{q^*_{\sigma_1}}{\sigma_1})}_{\text{part a}} + \underbrace{2(q^*_{\sigma_1} - q^*_{\sigma_2})}_{\text{part b}} \end{aligned} \tag{3.A.27}$$

For simplicity, denote part a in (3.A.27) as $j(x)$. Clearly $j(x) < 0$ for any $x \in \mathbb{R}$ with $j(x) \to 0$ as $q \to +\infty$. Note that part b is always positive, thus it only remains to show that there exists $\varepsilon > 0$, such that $j(x)$ is monotonically increasing for every $x > x_0 + \varepsilon$. The latter is valid, since the concavity of F ensures that both $F(x + \frac{q^*_{\sigma_1}}{\sigma_1}) - F(x + \frac{q^*_{\sigma_2}}{\sigma_2})$ and $F(x - \frac{q^*_{\sigma_2}}{\sigma_2}) - F(x - \frac{q^*_{\sigma_1}}{\sigma_1})$ are increasing for every $x > x_0 + \frac{q^*_{\sigma_2}}{\sigma_2}$, which completes the proof. □

Proof of Theorem 3.36. For every $\omega \in \Omega$, (3.A.1c) suggests that

$$\mathcal{H}_n(\sigma,\omega) \le \rho_D(F_\sigma, F_{n,\sigma}(\omega)) \le \mathcal{H}_n(\sigma,\omega) + h_\sigma(\mathcal{H}_n(\sigma,\omega)) \tag{3.A.28}$$

where $h_\sigma(x) = \min\{\sup_t |F_\sigma(x+t) - F_\sigma(t)|, \sup_t |F_{n,\sigma}(x+t) - F_{n,\sigma}(t)|\}$. For any $x \in [0,1]$, it is not difficult to show that

$$\lim_{\sigma\to\infty} h_\sigma(x) = 0.$$

Based on the definition of ρ_D, it is not difficult to see that for any $\sigma > 0$, $\rho_D(F_\sigma, F_{n,\sigma}(\omega)) = \rho_D(F_n(\omega), F)$. The result follows by taking the limit inferior on both sides of the first inequality in (3.A.28) and taking the limit superior on both sides of the second inequality in (3.A.28). □

Equation (3.3.12) of Theorem 3.36 has also been obtained by Rachev (1984) (see (2.2) therein).

Proof of Theorem 3.38. Equation (3.3.13) can be directly obtained from (3.2.14) and (2.3) of Rachev (1984), which is established with respect to the scaled Lévy metric. □

Proof of Theorem 3.39. Based on (3.3.8), when the sample comes from the alternative $G(x;n)$, we have

$$\begin{aligned}
\pi_n(\sigma) &= 1 - \mathbb{P}(\tilde{F}_-(x;\sigma,p) \le F_n(x) \le \tilde{F}_+(x;\sigma,p) \text{ for all } x) \\
&= 1 - \mathbb{P}((\tilde{F}_-(G^{-1}(t);\sigma,p)) \le \frac{1}{n}\eta_n(t) \le \tilde{F}_+(G^{-1}(t);\sigma,p) \text{ for all } 0<t<1) \\
&= 1 - \mathbb{P}(\sqrt{n}(\tilde{F}_-(G^{-1}(t);\sigma,p)) - \sqrt{n}t \\
&\qquad \le \sqrt{n}(\frac{1}{n}\eta_n(t) - t) \le \sqrt{n}\tilde{F}_+(G^{-1}(t);\sigma,p) - \sqrt{n}t, \text{ for all } 0<t<1),
\end{aligned} \tag{3.A.29}$$

where $\eta_n(t)$ is the process defined in (3.A.12). Substituting (3.3.7) and (3.4.1) in (3.A.29), we have

$$\begin{aligned}
\pi_n(\sigma) = 1 - \mathbb{P}\Big(&\sqrt{n}[F(F^{-1}(t) + \frac{1}{\sqrt{n}}\delta(t) - \frac{q^*_\sigma}{\sigma}) - q^*_\sigma - t] \le \sqrt{n}(\frac{1}{n}\eta_n(t) - t) \\
&\le \sqrt{n}[F(F^{-1}(t) + \frac{1}{\sqrt{n}}\delta(t) + \frac{q^*_\sigma}{\sigma}) + q^*_\sigma - t], \text{ for all } 0<t<1\Big)
\end{aligned} \tag{3.A.30}$$

On the one hand, since $\delta(t)$ is continuous and bounded, we have

$$\begin{aligned}\sqrt{n}\left[F(F^{-1}(t)+\frac{1}{\sqrt{n}}\delta(t)-\frac{q_\sigma^*}{\sigma})-F(F^{-1}(t)-\frac{q_\sigma^*}{\sigma})\right]\\ \sqrt{n}\left[F(F^{-1}(t)+\frac{1}{\sqrt{n}}\delta(t)+\frac{q_\sigma^*}{\sigma})-F(F^{-1}(t)+\frac{q_\sigma^*}{\sigma})\right]\end{aligned} \to \delta(t)f(F^{-1}(t))=\zeta(t),$$

$$\sqrt{n}(\frac{1}{n}\eta_n(t)-t) \xrightarrow{w} \mathbb{B}_0(t) \text{ on } L^\infty[0,1], \tag{3.A.31}$$

where the former convergences are uniform. On the other hand, from (3.3.6), (3.A.16), and the convergence in Theorem 3.21, we have

$$\sqrt{n}\left[F(F^{-1}(t)\pm\frac{q_\sigma^*}{\sigma})-t\pm q_\sigma^*\right] \to \pm\lambda_p(\sigma)(1+\frac{1}{\sigma}f(F^{-1}(t)))=\pm\gamma_\sigma(t), \tag{3.A.32}$$

uniformly. The required convergence relation (3.4.2) follows by substituting (3.A.31) and (3.A.32) in (3.A.30) and following the reasoning as in the proof of Theorem 3.21. □

Proof of Corollary 3.40. The result follows from Corollary 3.37, which states that the power $\pi_n(\infty)$ of the KS statistic is the limit of the power $\pi_n(\sigma)$ of $\mathcal{H}_n(\sigma)$ as $\sigma\to\infty$. □

Proof of Theorem 3.41. According to Theorem 7.2 in Bahadur (1971), in order to prove Theorem 3.41, we only need to prove that

$$\begin{aligned}H(F_\sigma,G_\sigma)&=\lim_{n\to\infty}\mathcal{H}_n(\sigma) \quad a.s.,\\ -f_\sigma(q)&=\lim_{n\to\infty}n^{-1}\log(1-L_{n,\sigma}(q)) \quad a.s.\end{aligned} \tag{3.A.33}$$

under the null hypothesis G.

To prove the first equation in (3.A.33), based on the triangular inequality, we obtain

$$H(F_\sigma,G_\sigma)-H(G_\sigma,F_{n,\sigma})\le\mathcal{H}_n(\sigma)\le H(F_\sigma,G_\sigma)+H(G_\sigma,F_{n,\sigma}). \tag{3.A.34}$$

Since

$$0\le H(G_\sigma,F_{n,\sigma})\le\rho_D(G_\sigma,F_{n,\sigma}),$$

$\rho_D(G_\sigma,F_{n,\sigma})\to 0$ a.s., under the alternative. The first equality in (3.A.33) follows

by taking $n \to \infty$ in (3.A.34).

In order to prove the second equality in (3.A.33), we apply Theorem 1 in Abrahamson (1967), then it is not difficult to show that

$$\begin{aligned}\mathbb{P}\{\mathcal{H}_n(\sigma) > q\} &\geq \mathbb{P}\{E_n(t) - t \geq F(F^{-1}(t) + \frac{q}{\sigma}) + q - t\},\\ \mathbb{P}\{\mathcal{H}_n(\sigma) > q\} &\geq \mathbb{P}\{E_n(t) - t \leq F(F^{-1}(t) - \frac{q}{\sigma}) - q - t\},\end{aligned} \tag{3.A.35}$$

where E_n is the empirical cdf of the sample $U_1, \ldots, U_n$ of n i.i.d uniform $(0,1)$ random variables.

Given a fixed t, we have,

$$\begin{aligned}\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(E_n(t) - t \geq F(F^{-1}(t) + \frac{q}{\sigma}) + q - t) &= -r(t, q + F(F^{-1}(t) + \frac{q}{\sigma}) - t),\\ \lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(E_n(t) - t \leq F(F^{-1}(t) - \frac{q}{\sigma}) - q - t) &= -r(1-t, q - F(F^{-1}(t) - \frac{q}{\sigma}) + t),\end{aligned} \tag{3.A.36}$$

which are obtained by substituting ϵ in (3.7a) and (3.7b) in Abrahamson (1967) with $q + F(F^{-1}(t) + \frac{q}{\sigma}) - t$ and $q - F(F^{-1}(t) - \frac{q}{\sigma}) + t$ respectively. Based on (3.A.35) and (3.A.36), we have

$$\begin{aligned}&\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\mathcal{H}_n(\sigma) > q)\\ &\geq \sup_{0<t<1} \max\{-r(t, q + F(F^{-1}(t) + \frac{q}{\sigma}) - t), -r(1-t, q - F(F^{-1}(t) - \frac{q}{\sigma}) + t)\} = -f_\sigma(q).\end{aligned} \tag{3.A.37}$$

On the other hand,

$$\begin{aligned}\{\mathcal{H}_n(\sigma) > q\} \subset &\bigcup_{0<t<1} \{E_n(t) - t \geq F(F^{-1}(t) + \frac{q}{\sigma}) + q - t\}\\ &\cup \bigcup_{0<t<1} \{E_n(t) - t \leq F(F^{-1}(t) - \frac{q}{\sigma}) - q - t\}.\end{aligned} \tag{3.A.38}$$

For any positive integer N, since $E_n(t) - t \leq E_n(\frac{i}{N}) - \frac{i-1}{N}$ when $\frac{i-1}{N} \leq t \leq \frac{i}{N}$, we have

$$\begin{aligned}&\bigcup_{0<t<1} \{E_n(t) - t \geq F(F^{-1}(t) + \frac{q}{\sigma}) + q - t\}\\ \subset &\bigcup_{i=1}^{N} \bigcup_{t \in [\frac{i-1}{N}, \frac{i}{N}]} \{E_n(t) - t \geq F(F^{-1}(t) + \frac{q}{\sigma}) + q - t\}\\ = &\bigcup_{i=1}^{N} \{E_n(\frac{i}{N}) - \frac{i}{N} \geq F(F^{-1}(\frac{i}{N}) + \frac{q}{\sigma}) + q - \frac{i+1}{N}\}.\end{aligned} \tag{3.A.39}$$

Therefore, for a sufficiently large N, taking probability on both sides of (3.A.39), we have

$$\begin{aligned}
&\mathbb{P}\left\{\bigcup_{0<t<1}\{E_n(t)-t\geq F(F^{-1}(t)+\frac{q}{\sigma})+q-t\}\right\}\\
\leq&\sum_{i=1}^{N}\mathbb{P}\{E_n(\frac{i}{N})-\frac{i}{N}\geq F(F^{-1}(\frac{i}{N})+\frac{q}{\sigma})+q-\frac{i+1}{N}\}\\
\leq&\sum_{i=1}^{N}\exp\{-nr(\frac{i}{N},q+F(F^{-1}(\frac{i}{N})+\frac{q}{\sigma})-\frac{i+1}{N})\}\\
\leq&N\sup_{0<t<1}\exp\{-nr(t,q+F(F^{-1}(t)+\frac{q}{\sigma})-t-\frac{1}{N})\}.
\end{aligned}\tag{3.A.40}$$

Then, from (3.A.38), (3.A.40) and an inequality similar to (3.A.40) with respect to $\bigcup_{0<t<1}\{E_n(t)-t\leq F(F^{-1}(t)-\frac{q}{\sigma})-q-t\}$, it follows that

$$\begin{aligned}
\mathbb{P}(\mathcal{H}_n(\sigma)>q)\leq 2N\sup_{0<t<1}\max\{&\exp\{-nr(t,q+F(F^{-1}(t)+\frac{q}{\sigma})-t-\frac{1}{N})\},\\
&\exp\{-nr(1-t,q-F(F^{-1}(t)+\frac{q}{\sigma})+t-\frac{1}{N})\}\}.
\end{aligned}\tag{3.A.41}$$

Taking first $\frac{1}{n}\log$ and then $\limsup$ as $n\to\infty$ on both sides of (3.A.41), we obtain

$$\begin{aligned}
&\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{P}(\mathcal{H}_n(\sigma)>q)\leq\\
&-\inf_{0<t<1}\min\{r(t,q+F(F^{-1}(t)+\frac{q}{\sigma})-t-\frac{1}{N}),r(1-t,q-F(F^{-1}(t)-\frac{q}{\sigma})+t-\frac{1}{N})\}.
\end{aligned}\tag{3.A.42}$$

Due to the arbitrariness of N in (3.A.42), one can choose $N=\infty$. Therefore the right hand side of (3.A.42) becomes equal to $-f_\sigma(q)$ as defined in (3.4.7), and from (3.A.42) we have

$$\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{P}(\mathcal{H}_n(\sigma)>q)\leq -f_\sigma(q).\tag{3.A.43}$$

The required second equality in (3.A.33) follows by combining (3.A.37) and (3.A.43).

□

3.B Power of the Scaled Hausdorff Statistic for Different σ

In Section 3.3, we have analysed theoretically how the choice of scale parameter σ affects the power $\pi_n(\sigma)$ of the $\mathcal{H}_n(\sigma)$ statistic. In this section, we verify numerically

Table 3.7: The Parameters of the Heavy-tailed G in Example 3.51

C	$F(C)=\phi_1$	Tail	$G_0(x)$	(α,θ)
$\log 2$	0.5	Pareto	$1-(\frac{\alpha}{x+\alpha})^{\theta}$	$(\frac{\log 2}{2^{1/3}-1},3)$
$\log 5$	0.8	Lognormal	$LN(\alpha,\theta)$	$(\log 5-1.2\Phi^{-1}(0.8),1.2^2)$
$\log 5$	0.8	Weibull	$1-e^{-(x/\alpha)^{\theta}}$	$((\log 5)^{1/2},2)$

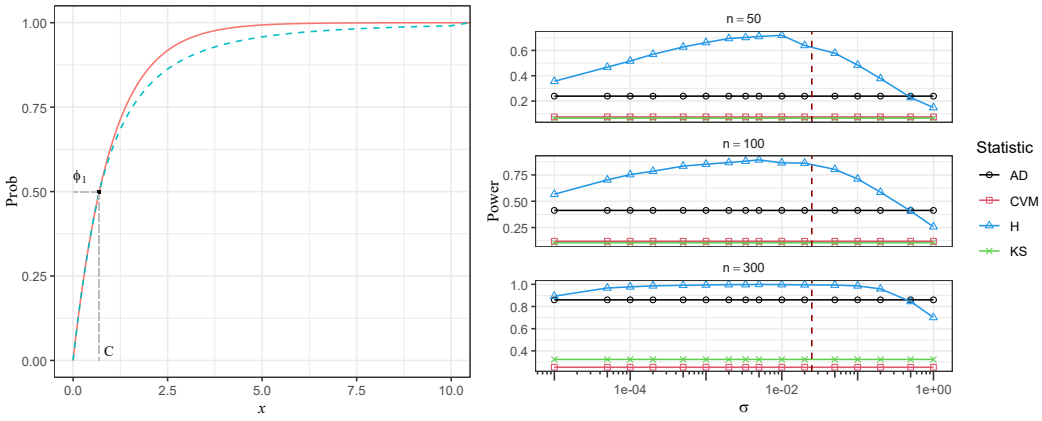


Figure 3.16: Left panel: $F\sim Exp(1)$ (solid) and G (dashed) is spliced with $C=\log 2$, $F(C)=\phi_1=0.5$ and a Pareto tail $G_0(x)=1-(\frac{\alpha}{x+\alpha})^{\theta}$ where $\theta=3$ and $\alpha=\frac{\log 2}{2^{1/3}-1}$; Right Panel: the powers of the $H_n(\sigma)$, KS, CvM and AD tests as a function of σ, with σ^* defined in (3.3.11) with $\psi=(0.99,0.95)$ and indicated by the dashed vertical darkred lines.

these theoretical considerations, by applying the splicing construction for the alternative $G(x)$ as in Example 3.43. The density $g(x)$ of $G(x)$ is defined as in (3.4.9) and is assumed to satisfy conditions 1)-3) of Example 3.43. In other words, G coincides with F in the body and has the tail of a preselected distribution G_0.

Example 3.51 (Light-tail Null vs Heavy-tail Sample)**.** We take $F\sim Exp(1)$, $\underline{x}=0$, and G to be a spliced distribution with body $G(x)=F(x)$, $x\in[\underline{x},C]$ and tail $G(x)=G_0(x)$, $x>C$, where G_0 is taken to be Pareto, Lognormal, and Weibull, with splicing point C and weight ϕ_1 specified in Table 3.7, satisfying condition 1)-3) in Example 3.43, where Φ^{-1} denotes the quantile function of standard Normal. The shapes of the null F and the alternative G from Table 3.7 are illustrated graphically in the left panels of Figures 3.16, 3.17 and 3.18.

The right panels of Figures 3.16, 3.17 and 3.18 illustrate the power $\pi_n(\sigma)$ of $\mathcal{H}_n(\sigma)$, compared with the powers of KS, AD and CvM, for different sample sizes $n=$50, 100 and 300 and different choices of the scale parameter σ, ranging from 10^{-5} to 1. As can be seen, the p-values (and therefore the powers) of KS, AD and

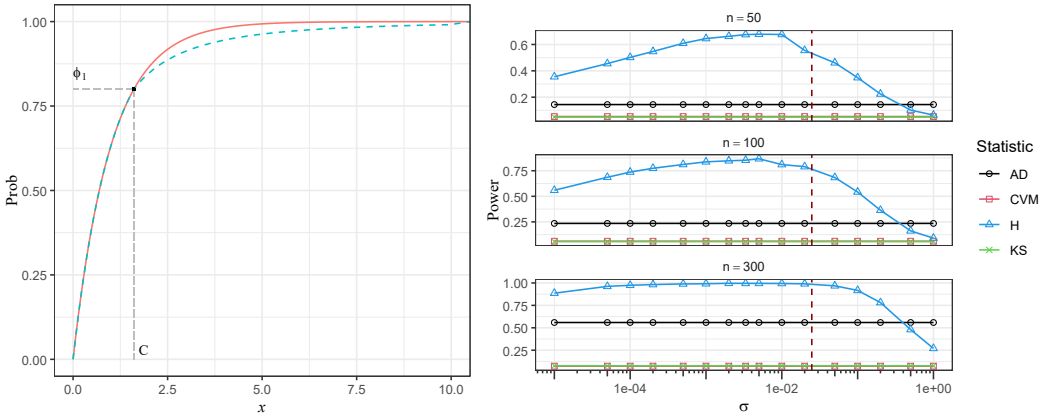


Figure 3.17: Left panel: $F \sim Exp(1)$ (solid) and G (dashed) with $C = \log 5$, $F(C) = \phi_1 = 0.8$ and a Lognormal tail $G_0 \sim LN(\alpha, \theta)$ where $\theta = 1.2^2$ and $\alpha = \log 5 - 1.2\Phi^{-1}(0.8)$; Right Panel: the powers of the $H_n(\sigma)$, KS, CvM and AD tests as a function of σ, with σ^* defined in (3.3.11) with $\boldsymbol{\psi} = (0.99, 0.95)$ shown by the dashed vertical darkred lines.

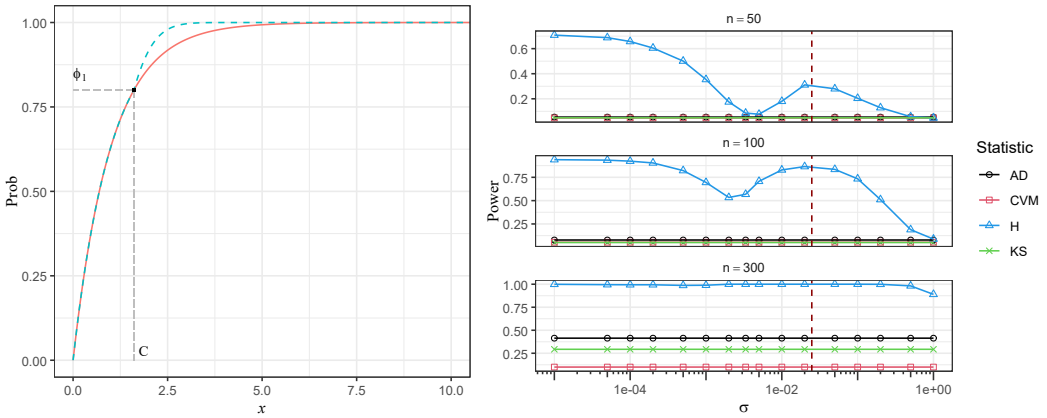


Figure 3.18: Left panel: $F \sim Exp(1)$ (solid) and G (dashed) is spliced with $C = \log 5$, $F(C) = \phi_1 = 0.8$ and a Weibull tail $G_0(x) = 1 - e^{-(x/\alpha)^\theta}$ where θ =2 and $\alpha = (\log 5)^{1/2}$; Right Panel: the powers of the $H_n(\sigma)$, KS, CvM and AD tests as a function of σ, with σ^* defined in (3.3.11) with $\boldsymbol{\psi} = (0.99, 0.95)$ shown by the dashed vertical darkred lines.

CvM are invariant to σ. The "optimal" value $\sigma^* = \frac{0.99-0.95}{\ln 0.05 - \ln 0.01}$ suggested by the rule (3.3.11) with $\boldsymbol{\psi} = (0.99, 0.95)$ is indicated by dashed vertical darkred line in the right panel of Figure 3.16. As shown, when the null F is light-tailed and the alternative G is heavy-tailed, the dashed red line corresponding to σ^* approximates reasonably well the local maxima of $\pi_n(\sigma)$, which are also global maxima in the case of Pareto and Lognormal tail (see the right panels of Figures 3.16 and 3.17). In the case of Weibull tail, the global maximum of $\pi_n(\sigma)$ is achieved at $\sigma \to 0$, which is degenerate and makes no practical sense, since transforming the scale with a very

Table 3.8: The Parameters of Heavy-tailed F and Light-tailed G in Example 3.52

Null F				Alternative G	
	CDF	$\underline{x}$	(α,θ)	C	$F(C)=\phi_1$
Pareto	$F(x)=1-\frac{\alpha^\theta}{(x+\alpha)^\theta}$	0	$(2,2.5)$	0.8853998	0.6
Lognormal	$LN(\alpha,\theta)$	0	$(0,1.2^2)$	2.745451	0.8
Weibull	$F(x)=1-e^{-(x/\alpha)^\theta}$	0	$(1,0.5)$	2.590290	0.8

small σ zeroises the entire sample $\sigma \boldsymbol{X}_n$.

Example 3.51 validates numerically the theoretical results of Section 3.3. We have also shown that by choosing $\sigma=\sigma^*$ following (3.3.11), the power of $\mathcal{H}_n(\sigma^*)$ becomes significantly higher than the powers of AD, CvM and KS (as seen from Figures 3.16, 3.17 and 3.18).

In Example 3.52, we investigate the power $\pi_n(\sigma)$ of $\mathcal{H}_n(\sigma)$ as a function of σ when F is heavy-tailed and G is light-tailed. More precisely, we assume that F is a heavy-tailed distribution and G is a spliced distribution as in Example 3.43 with a body coincident with F and a light exponential tail. For the heavy-tailed null distribution F, we consider three cases, Pareto, Weibull, and Lognormal, summarized in Example 3.52 with parameters in Table 3.8.

Example 3.52 (Heavy-tail Null vs Light-tail Sample)**.** The null cdfs are taken to be Pareto Weibull, or Lognormal, and the alternative G to be a spliced distribution with body $G(x)=F(x)$, $x\in[\underline{x},C]$ and tail $G(x)=G_0(x)$, $x>C$, $G_0\sim Exp(\alpha_e)$ with $\alpha_e=-\log(1-\phi_1)/C$ so that condition 3) of Example 3.43 is met. The parameters of the null F and values of splicing point C and weight ϕ_1 are summarized in Table 3.8.

The shapes of F and G from Table 3.8 are illustrated in the left panels of Figures 3.19, 3.20 and 3.21 where in the right panels, we illustrate the powers of KS, AD, CvM and $\mathcal{H}_n(\sigma)$, for sample sizes $n=50$, 100 and 300 and different choices of σ, ranging from 10^{-5} to 1. In the latter right pannes, the corresponding "optimal" value σ^* are indicated by the dashed vertical darkred lines. As can be seen, σ^* is reasonably close to the local maximum of $\mathcal{H}_n(\sigma)$, where the approximation improves when the sample size n increases.

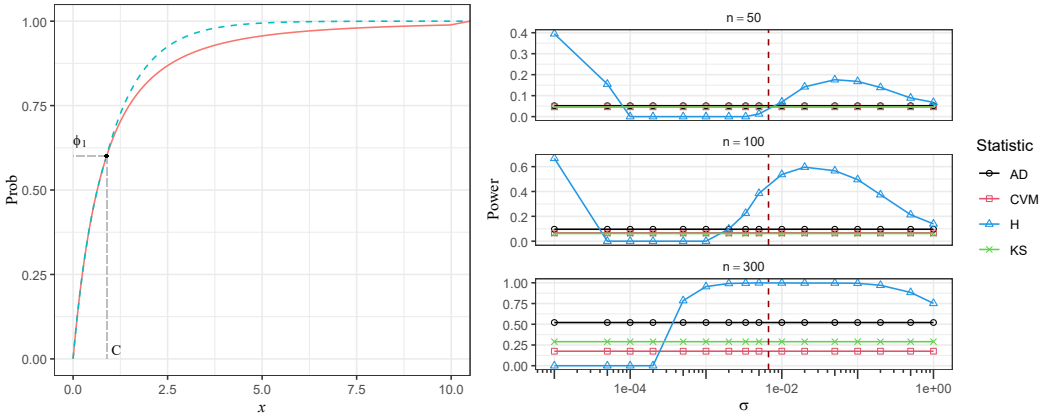


Figure 3.19: Left panel: $F(x)$ (solid) is Pareto with $\alpha = 2$ and $\theta = 2.5$ and G (dashed) is spliced with $C = 0.8853998$, $F(C) = \phi_1 = 0.6$ and an Exponential tail $G_0 \sim Exp(\alpha_e)$ where $\alpha_e = -\log(1-\phi_1)/C$; Right Panel: the powers of the $H_n(\sigma)$, KS, CvM and AD tests as a function of σ, with σ^* defined in (3.3.11) with $\boldsymbol{\psi} = (0.99, 0.95)$ shown by the dashed vertical darkred lines.

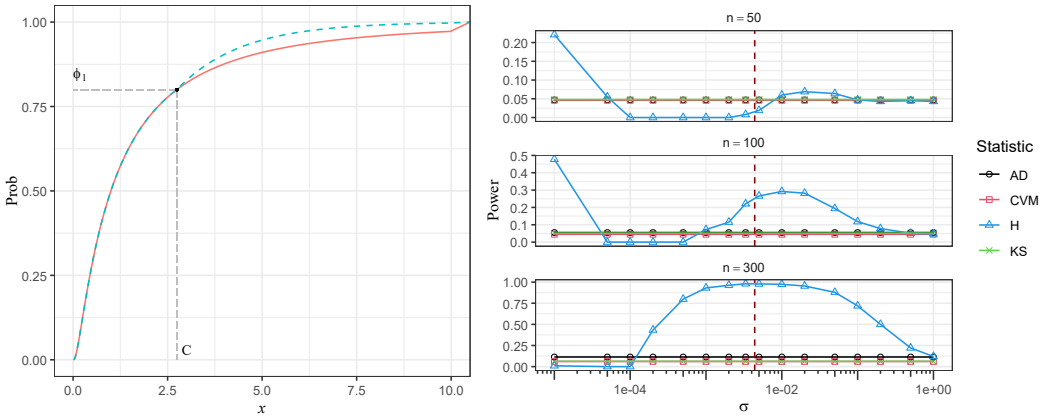


Figure 3.20: Left panel: $F \sim LN(\alpha, \theta)$ (solid) to be Lognormal with $\alpha = 0$ and $\theta = 1.2^2$ and G (dashed) is spliced with the splicing point $C = 2.745451$, $F(C) = \phi_1 = 0.8$ and an Exponential tail $G_0 \sim Exp(\alpha_e)$ where $\alpha_e = -\log(1-\phi_1)/C$; Right Panel: the powers of the $H_n(\sigma)$, KS, CvM and AD tests as a function of σ, with σ^* defined in (3.3.11) with $\boldsymbol{\psi} = (0.99, 0.95)$ shown by the dashed vertical darkred lines.

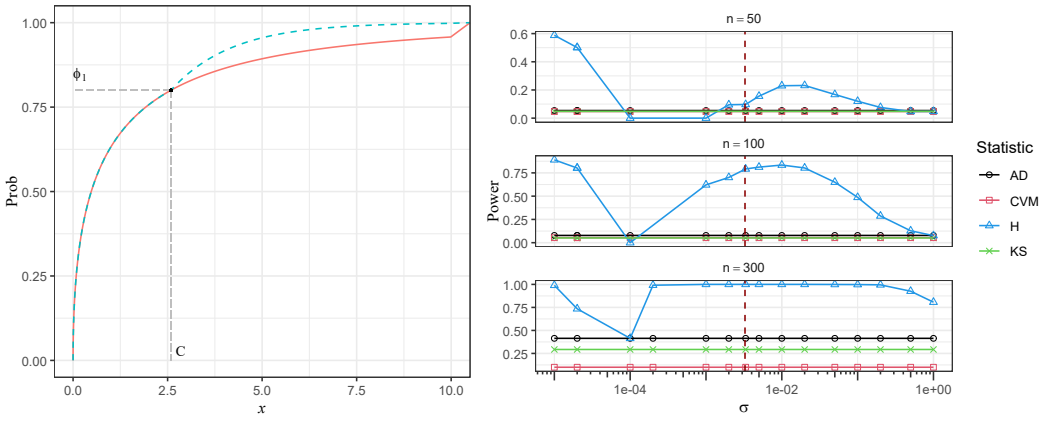


Figure 3.21: Left panel: $F(x) = 1 - e^{-(x/\alpha)^\theta}$ (solid) to be Weibull distribution with $\alpha = 1$ and $\theta = 0.5$ and G (dashed) is spliced with the splicing point $C = 2.59029$, $F(C) = \phi_1 = 0.8$ and an Exponential tail $G_0 \sim Exp(\alpha_e)$ where $\alpha_e = -\log(1 - \phi_1)/C$; Right Panel: the powers of the $H_n(\sigma)$, KS, CvM and AD tests as a function of σ, with σ^* defined in (3.3.11) with $\boldsymbol{\psi} = (0.99, 0.95)$ shown by the dashed vertical darkred lines.

Chapter 4

On a Two-sample Multivariate Goodness-of-fit Test based on the Hausdorff Metric

This chapter is based on the paper:

Dimitrina S. Dimitrova, Yun Jia and Vladimir K. Kaishev (2025). On a Two-sample Multivariate Goodness-of-fit Test based on the Hausdorff Metric. *near submission.*

Abstract

We introduce a two-sample goodness-of-fit test based on the Hausdorff distance between possibly multivariate empirical cumulative distribution functions (ecdfs). Classical tests such as Kolmogorov–Smirnov (KS), Cramér–von Mises (CvM), and Anderson–Darling (AD) have appealing properties in the univariate setting but lose efficiency for tail differences. Existing multivariate tests, including Wasserstein and run-based tests, can achieve high power but are computationally demanding. Since no hypothesized null distribution is available, extending the existing approach in Chapter 3 to the two-sample multivariate setting is challenging. We address these challenges as follows. We introduce an explicit and computable representation of the two-sample Hausdorff (H) statistic with a geometric interpretation as the edge of the largest hypercube that can be inscribed between the two ecdfs. We propose a permutation version of H and establish its asymptotic equivalences in terms of power and type I error, under the null and (fixed or contiguous) alternative. Based on this, we develop a method to compute the exact and asymptotic p-values of the

H statistic.

We address these challenges as follows. We introduce an explicit and computable representation of the two-sample Hausdorff (H) statistic with a geometric interpretation as the edge of the largest hypercube that can be inscribed between the two ecdfs. We propose a permutation version of H and establish its asymptotic equivalences in terms of power and type I error, under the null and (fixed or contiguous) alternative. Based on this, we develop a method to compute the exact and asymptotic p-values of the H statistic. In view of the scale dependence of H, we propose a rule for selecting the scale coefficient, so as to optimize its power. Last but not least, we give some useful properties of H including its Lipschitz continuity, qualitative robustness and connections to the Lévy-Prokhorov metric and the KS test. We demonstrate based on numerical examples that the scale-tuned Hausdorff test outperforms the major competitors in terms of power in the univariate and bivariate cases.

4.1 Introduction

Consider the two-sample goodness-of-fit problem of whether two random samples come from one and the same unknown multivariate distribution. The classical goodness-of-fit tests in the case when the two samples are drawn from univariate distributions, include the Kolmogorov-Smirnov (KS), the Kuiper, the Cramer von Mises (CvM), and the Anderson-Darling (AD), introduced correspondingly, by Kolmogorov (1933), Smirnov (1939), Kuiper (1960), Cramér (1928), Von Mises (1931) and Anderson and Darling (1952). The latter tests have gained significant popularity and have been widely applied in almost any field where data is collected and analysed, such as, astronomy (McQuillan et al., 2013), social sciences (Salman et al., 2015), pattern recognition (Alzubaidi and Kalita, 2016), machine learning (Gretton et al., 2012) etc., to name only a few. These tests are based on distances between empirical distribution functions, which are easy to compute in the univariate case. Due to their popularity, some of these tests are further extended to the multivariate case by e.g. Peacock (1983), Kim et al. (2020).

In the multivariate case, alternative definitions of distances between the samples and related goodness-of-fit statistics have also been considered, including the run tests based on the minimal spanning tree due to Friedman and Rafsky (1979), and on

the shortest Hamiltonian path proposed by Biswas et al. (2014), the Wasserstein test (Hundrieser et al., 2024), the Ball divergence test (Pan et al., 2018), the Maximum Mean Discrepancy test (Gretton et al., 2012), the Cross Match test (Rosenbaum, 2005) and the Schilling-Henze Nearest Neighbor test (cf. Schilling, 1986; Henze, 1988).

These and other existing two-sample tests with different properties are a popular tool for classification and unsupervised learning. However, there is no 'best', test that suits all purposes and possesses all the best properties. For example, the very popular KS test, based on the supremum distance, is readily understood graphically, is easy to evaluate, and is distribution-free in the univariate case, when the null is continuous. Furthermore, recently, Dimitrova et al. (2020) provided efficient means of computing the KS p-values assuming arbitrary, continuous, discrete, or mixed null distribution, which makes the KS test applicable beyond just the continuous case. At the same time, the KS statistic, is less sensitive in the tails, and has in general lower power (see e.g. Mason and Schuenemeyer, 1983; Feigelson and Babu, 2020). This makes the KS test less efficient, especially for comparing tails, which is very important in extreme value applications and related inference. The Wasserstein test and the run tests in the multidimensional setting also have high power but are difficult to evaluate numerically. The evaluation of the latter run statistics may be a non-deterministic polynomial time (NP) problem, which hinders their practical use.

The definition of the distance between the two samples determines the properties of the test, in particular the evaluation of the test, its p-values and power. All these considerations lead to the conclusion that there is still scope for using alternative distance metrics, leading to the construction of new test statistics, and the need to investigate their related properties.

The aim of this chapter is to explore how the Hausdorff metric, introduced by Hausdorff (1914), to measure the distance between sets, can be applied to measure the distance between multivariate empirical cumulative distribution functions (ecdfs), with the purpose of introducing a corresponding two-sample Hausdorff goodness-of-fit test statistic. The Hausdorff distance has been considered by Beer (1984); Sendov and Beer (2012), within the context of approximation theory, and more recently in machine learning, by Chavent (2004), Li et al. (2017), Karimi and

Salcudean (2020), Zhao et al. (2021), to name only a few of the papers in this stream of literature. In Chapter 3, we have proposed to apply the Hausdorff (H) distance in the context of one-sample univariate goodness-of-fit testing. By investigating theoretically the scale dependence of the H test and its p-values, we have shown that its power can be controlled and optimized by appropriately selecting the scale. As they note, knowledge of the hypothesized null distribution is central in implementing the latter power optimization. As demonstrated, this leads to the H test significantly outperforming the classical KS, CvM and AD tests in terms of statistical power. Efficient numerical methods to compute the Hausdorff metric, the exact and asymptotic p-values and the asymptotic power and Bahadur efficiency of the related H-test have also been provided.

However, extending these results to the two-sample case is not straightforward since, no null distribution is hypothesized, in contrast to the one-sample case. Furthermore, the methods to compute the one-sample Hausdorff statistic and its p-values assume continuity of the null distribution, which is not defined in the two-sample case, where H measures the distance between two empirical cdfs. Evaluating numerically a distance measure between possibly multivariate ecdfs represents yet another challenge by itself.

In this paper, we address all these challenges. Our major contributions can be summarized as follows. First, we introduce the two-sample Hausdorff (H) test as the distance between two possibly multivariate empirical cdfs. In Theorem 4.5 we show that the H test can be interpreted geometrically as the edge of the maximum hypercube that can be inscribed between the two ecdfs. Based on Theorem 4.5, we prove Lemma 4.15, which gives an explicit and computationally appealing expression of the two-sample multivariate H test. Theorem 4.5 and Lemma 4.15 are generalizations of Lemmas 3.8 and 3.9 derived in Section 3.2 for the case of one-sample drawn from a univariate distribution. In Section 4.3.1 and Appendix 4.A, we give methods to compute the H statistic when the two samples come from univariate and bivariate distributions.

Second, we develop a method to compute the p-value of the H test by introducing an appropriate permutation H test in (4.3.10). This is necessary since the p-value of the permutation H test is based on the observable pooled sample, in

contrast to the p-value of the H test, which depends on the unknown null. This is further justified by relation (4.3.11), Theorems 4.24 and 4.25, where we show that the discrepancy between the p-values of the H test and its permutation version (4.3.10) is asymptotically negligible under the null, or under fixed and contiguous alternatives. Then in Theorem 4.26, we show that the univariate permutation H test controls the type I error under the null, and also yields the same power as the original test under the fixed and contiguous alternative. In addition, in Theorem 4.28, we also give an expression for the asymptotic p-value of the H statistic.

Third, in Section 4.4, we address the problem of scale dependence of the H test, its p-values and power, and provide a rule to select the scale so as to optimize the power. Rule (3.3.11) proposed in Section 3.3.2 for the one-sample statistic is not directly applicable, since no distributional assumptions are made in two-sample goodness of fit testing. Alternatively, based on the permutation H test, we propose the rule (4.4.6) and its multivariate generalization (4.4.9) to tune the two-sample statistic H in the spirit of (3.3.11). We further show that rule (4.4.6) is asymptotically equivalent to the one-sample rule (3.3.11). When samples deviate from each other in the right tail, we demonstrate by numerical example that in the univariate case, the power of scale-tuned H test is higher than that of the KS, CvM, and AD tests, and is very close to the power of the Wasserstein test when sample sizes are moderately large (see Figure 4.8). In the bivariate case, we show by numerical comparison that the power of the H test is higher than that of other multidimensional tests and is significantly higher than the power of the KS test generalized by Peacock (1983) (see Table 4.1). In Theorem 4.36, we also give an expression of the exact Bahadur slope of the H test.

Last but not least, we provide some useful properties of the H statistic. In Theorem 4.6, we establish its Lipschitz continuity as a point function of the sample observations which we combine with Theorem 4.25 to establish the qualitative robustness of H. In Proposition 4.9, we show that H can be viewed as a degenerate Lévy-Prokhorov metric. Theorem 4.19 establishes how the two-sample univariate H test is related to the two sample KS test.

This paper is organized as follows. In Section 4.2, we recall the general definition of the Hausdorff metric and the empirical cdf in the multivariate case, and propose

the Hausdorff statistic to be the distance between empirical cdfs. We further provide some useful properties of the H statistic among which its connection to the KS statistic and the Lévy-Prokhorov distance. In Section 4.3, we give detailed methods to compute the H statistic and its p-values in the one- and two-dimensional cases. In Section 4.4, we show how the p-values of the H test depend on the scale and provide a rule to select the scale so as to optimize its power. The latter rule is illustrated by numerical examples. Finally, in Section 4.6, we summarize our findings.

In Appendix 4.A, we give a detailed method to compute the H statistic in the bivariate case. In Appendix 4.B, we provide an expression of the H p-values as a boundary crossing problem. The proofs of all the results that appear in the paper are given in Appendix 4.C. In Appendix 4.D, we compare the numerical efficiency of the two methods in Section 4.3.1.

4.2 The Hausdorff Goodness-of-fit Test Statistic

Let two samples $\boldsymbol{X}_m = \{X_1, \ldots, X_m\}$ and $\boldsymbol{Y}_n = \{Y_1, \ldots, Y_n\}$ be defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where X_i and Y_j, $i = 1, \ldots, m$, $j = 1, \ldots, n$ are i.i.d. copies of the k-dimensional random vectors X and Y, i.e.

$$X_i = (X_i^{(1)}, \ldots, X_i^{(k)}),\ Y_j = (Y_j^{(1)}, \ldots, Y_j^{(k)}).$$

Denote by $F(x)$ and $G(x)$ the (unknown) cdfs of the the random vectors X and Y respectively. We assume that $F(x)$ and $G(x)$ as arbitrary distribution functions, i.e. repeated observations are allowed in the two samples. We want to test the null hypothesis $H_0 : F(x) = G(x)$, for all $x \in \mathbb{R}^k$, against the alternative $H_1 : F(x) \neq G(x)$ for at least one $x \in \mathbb{R}^k$. We are interested in applying the Hausdorff metric to measure the distance between two empirical distribution functions for goodness-of-fit testing.

4.2.1 Empirical Distribution Funtions

When X and Y are univariate random variables, i.e. $k = 1$, denote by $F_m(x)$, $G_n(x)$ and $E_{n+m}(x)$ $(x \in \mathbb{R})$ the empirical cumulative distribution functions (ecdfs) corresponding to the samples $\boldsymbol{X}_m, \boldsymbol{Y}_n$ and the pooled sample $\boldsymbol{Z}_{m+n} =$

$\{X_1,\ldots,X_m,Y_1,\ldots,Y_n\}=\{Z_1,\ldots,Z_{m+n}\}$, i.e.

$$F_m(x)=\frac{1}{m}\sum_{j=1}^{m}\mathbb{1}(X_j\le x),\ G_n(x)=\frac{1}{n}\sum_{j=1}^{n}\mathbb{1}(Y_j\le x),\ E_{m+n}(x)=\frac{1}{m+n}\sum_{j=1}^{m+n}\mathbb{1}(Z_j\le x), \tag{4.2.1}$$

where $x\in\mathbb{R}$ and $\mathbb{1}(\cdot)$ is the indicator function.

When X and Y are multivarite, i.e. $k\ge 2$, it is well known that the ecdfs of samples have multiple alternative definitions, (see e.g. Langrené and Warin, 2021). Hence to define the latter, we need to first introduce the component-wise order $\preceq$ on $\mathbb{R}^k$. For any $x_1,x_2\in\mathbb{R}^k$, we say $x_1\preceq x_2$, if and only if $x_1^{(i)}\le x_2^{(i)}$, $i=1,\ldots,k$, where $x_1^{(i)}$ and $x_2^{(i)}$ are the i-th components of x_1 and x_2. Given the component-wise order, the corresponding empirical distribution functions of $\boldsymbol{X}_m$ and $\boldsymbol{Y}_n$ is

$$F_m(x)=\frac{1}{m}\sum_{j=1}^{m}\mathbb{1}(X_j\preceq x),\ G_n=\frac{1}{n}\sum_{j=1}^{n}\mathbb{1}(Y_j\preceq x)\ x\in\mathbb{R}^k. \tag{4.2.2}$$

However, one can have 2^k-1 alternative component-wise order definitions on $\mathbb{R}^k$, which we denote as $\preceq_i$ $(i=2,\ldots,2^k)$. For example, when $k=2$, the alternative component-wise order definitions are

$$\begin{aligned}
&x_1\preceq_2 x_2 \text{ if and only if } x_1^{(1)}\le x_2^{(1)} \text{ and } x_1^{(2)}\ge x_2^{(2)},\\
&x_1\preceq_3 x_2 \text{ if and only if } x_1^{(1)}\ge x_2^{(1)} \text{ and } x_1^{(2)}\le x_2^{(2)},\\
&x_1\preceq_4 x_2 \text{ if and only if } x_1^{(1)}\ge x_2^{(1)} \text{ and } x_1^{(2)}\ge x_2^{(2)}.
\end{aligned} \tag{4.2.3}$$

We will denote by $F_{m,i}(x)$ the empirical cdf defined by the i-th component-wise order $\preceq_i$. Clearly $\preceq_1\equiv\preceq$, therefore $F_m\equiv F_{m,1}$ and $G_n\equiv G_{n,1}$.

In the sequel, it will be convenient to interchangeably use the notation $F_m(x)\equiv F_m(x,\omega)\equiv F_m$, $G_n(x)\equiv G_n(x,\omega)\equiv G_n$ and $E_{m+n}(x)\equiv E_{m+n}(x,\omega)\equiv E_{m+n}$, where $\omega\in\Omega$ explicitly indicates that the ecdfs are random realizations underpinned by correspondingly the random samples $\boldsymbol{X}_m(\omega)$, $\boldsymbol{Y}_n(\omega)$ and $\boldsymbol{Z}_{m+n}(\omega)$.

4.2.2 Background on the Hausdorff Metric

In Chapter 3, we have considered the Hausdorff metric and its use to measure the distance between a continuous cdf and an empirical cdf in the one-sample one-dimensional case. In this section, we will extend the definitions of the Hausdorff metric and statistic in Section 3.2.1 to the two-sample multi-dimensional case. For

the purpose, we recall the Definition 3.1, which defines a distance measure between two points in $\mathbb{R}^{k+1}$.

Definition 4.1. Let $A, B \in \mathbb{R}^{k+1}$ with coordinates $x_A, x_B \in \mathbb{R}^k$, $z_A, z_B \in \mathbb{R}$, i.e. $A = (x_A, z_A)^T$, $B = (x_B, z_B)^T$, where $\cdot^T$ denotes transposition. The function $\rho(A, B)$ is a distance measure between two points $A, B \in \mathbb{R}^{k+1}$ iff it satisfies the conditions (1)-(4) in Definition 3.1, i.e.:

1. $\rho(A, B) \geq 0$, for every pair of points A and B.

2. $\rho(A, B) = 0$, iff $A = B$.

3. Symmetry: $\rho(A, B) = \rho(B, A)$.

4. Triangle inequality: $\rho(A, B) + \rho(B, C) \geq \rho(A, C)$, $C \in \mathbb{R}^{k+1}$.

For example, we could easily verify that the Chebyshev and Euclidean distances, defined by

$$\begin{aligned} \rho_\infty(A, B) &= \max_{1 \leq i \leq k+1} |w_A^{(i)} - w_B^{(i)}|, \\ \rho_2(A, B) &= \sqrt{\sum_{i=1}^{k+1} (w_A^{(i)} - w_B^{(i)})^2}, \end{aligned} \tag{4.2.4}$$

respectively, satisfy Definition 4.1, where $w_A^{(i)}$ and $w_B^{(i)}$ are the i-th components of points A and B respectively. Given an arbitrary measure $\rho(A, B)$ of distance between two points in $\mathbb{R}^{k+1}$, the Definition 3.3 of the Hausdorff distance can be generalized as follows.

Definition 4.2. Let $\mathcal{A}$ and $\mathcal{B}$ be two arbitrary sets in $\mathbb{R}^{k+1}$. The Hausdorff distance, $H_\rho(\mathcal{A}, \mathcal{B})$ between the sets $\mathcal{A}$ and $\mathcal{B}$ is defined as

$$H_\rho(\mathcal{A}, \mathcal{B}) = \max[\sup_{A \in \mathcal{A}} \inf_{B \in \mathcal{B}} \rho(A, B), \sup_{B \in \mathcal{B}} \inf_{A \in \mathcal{A}} \rho(A, B)]. \tag{4.2.5}$$

In order to apply the Hausdorff metric $H_\rho(\mathcal{A}, \mathcal{B})$ to measure the distance between two ecdfs, $F_m(x, \omega), G_n(x, \omega)$, $x \in \mathbb{R}$, we need to appropriately define the sets $\mathcal{A}$ and $\mathcal{B}$ from Definition 4.2, replacing them with the corresponding planar curve (set) analogs of $F_m(x, \omega)$ and $G_n(x, \omega)$.

The latter planar curve analogues $F_m^c(\omega)$ and $G_n^c(\omega)$ in the univariate case, i.e. for $k=1$ are defined as

$$\begin{aligned} F_m^c(\omega) &= \{(x,z) : F_m(x-,\omega) \le z \le F_m(x+,\omega)\} \\ G_n^c(\omega) &= \{(x,z) : G_n(x-,\omega) \le z \le G_n(x+,\omega)\}, \end{aligned} \tag{4.2.6}$$

where $F(x-)$ and $F(x+)$ are the left and right limits at the point $x \in \mathbb{R}$. When $k>1$, we need to adequately define $F_m(x-)$, $F_m(x+)$, $G_n(x-)$ and $G_n(x+)$ for $x \in \mathbb{R}^k$ $(k>1)$. This task has been considered by Popov (1999), who refers to these limits as lower and upper Baire functions. For example, the Baire functions of the ecdfs $F_{m,i}$, are

$$\begin{aligned} F_{m,i}(x-) &= \liminf_{t \to x} F_{m,i}(t), \\ F_{m,i}(x+) &= \limsup_{t \to x} F_{m,i}(t). \end{aligned} \tag{4.2.7}$$

Similar definitions for $G_{n,i}$ are also adopted. Note that the planar curves F_m^c and G_n^c defined by (4.2.6) and (4.2.7) are the minimum closed and connected sets in $\mathbb{R}^{k+1}$ containing the graphs of the functions $F_m(x,\omega)$ and $G_n(x,\omega)$ (cf. Lemma 2 in Popov, 1999). The latter graphs are $gr(F_m(x,\omega)) = \{(x, F_m(x,\omega)), \forall x \in \mathbb{R}\}$, and $gr(G_n(x,\omega)) = \{(x, G_n(x,\omega)), \forall x \in \mathbb{R}\}$, respectively, which are unclosed and unconnected sets. These graphs and their corresponding planar curves are illustrated for the case $k=1$. For this purpose, we introduce the following notation. Denote by $(x_1,\ldots,x_\upsilon)$ and $(y_1,\ldots,y_\nu)$ the corresponding distinct values $x_1 < x_2 < \cdots < x_\upsilon$ and $y_1 < y_2 < \cdots < y_\nu$ of the observations $(X_1,\ldots,X_m)$ and $(Y_1,\ldots,Y_n)$, where each x_i and y_j is repeated r_i and e_j times, respectively, with $\sum_{i=1}^{\upsilon} r_i = m$ and $\sum_{j=1}^{\nu} e_j = n$. Clearly, for the ecdfs F_m and G_n, we have

$$\begin{aligned} F_m(x_1) = \frac{r_1}{m}; F_m(x_2) = \frac{r_1+r_2}{m}, \ldots, F_m(x_\upsilon) = \frac{r_1+\cdots+r_\upsilon}{m} \\ G_n(y_1) = \frac{e_1}{n}; G_n(y_2) = \frac{e_1+e_2}{n}, \ldots, G_n(y_\nu) = \frac{e_1+\cdots+e_\nu}{n}. \end{aligned}$$

Clearly, the planar curves F_m^c and G_n^c that correspond to the cdfs F_m and G_n represent piecewise linear curves, with horizontal and vertical linear segments forming steps. In order to define the latter staircases, we will introduce the notation, $x_1 = a_1$, $x_2 - x_1 = a_2$, $\ldots$, $x_\upsilon - x_{\upsilon-1} = a_\upsilon$ and $b_0 = 0$, $\frac{r_i}{m} = b_i$, $i = 1,\ldots,\upsilon$ and $c_0 = a_1$, $c_1 = y_1 - c_0$, $y_2 - y_1 = c_2$, $\ldots y_\nu - y_{\nu-1} = c_\nu$, $d_0 = 0$, $\frac{e_j}{m} = d_j$, $j = 1,\ldots,\nu$

and $a_{\upsilon+1} = \sum_{j=0}^{\nu} c_j - \sum_{j=1}^{\upsilon} a_j$. Without loss of generality, assume $a_1 < c_1$. Let F_m^c have vertices $A_1, A_2, \ldots, A_{2\upsilon}, A_{2\upsilon+1}$, with coordinates on a positively orientated orthogonal, Cartesian coordinate system xOz in $\mathbb{R}^2$, given as

$$A_{2l} = \left(\sum_{j=1}^{l} a_j, \sum_{j=1}^{l} b_j\right) \; l = 1,2,\ldots,\upsilon; \; A_{2l+1} = \left(\sum_{j=1}^{l+1} a_j, \sum_{j=0}^{l} b_j\right) \; l = 0,1,\ldots,\upsilon. \quad (4.2.8)$$

Similarly, let G_n^c have vertices $B_0, B_1, \ldots, B_{2\nu}$ with coordinates

$$B_{2l} = \left(\sum_{j=0}^{l} c_j, \sum_{j=0}^{l} d_j\right) \; l = 0,1,\ldots,\nu; \; B_{2l+1} = \left(\sum_{j=0}^{l+1} c_j, \sum_{j=0}^{l} d_j\right) \; l = 0,1,\ldots,\nu-1. \quad (4.2.9)$$

The stepwise curves F_m^c and G_n^c, may have arbitrary numbers of vertices $2\upsilon+1$ and $2\nu+1$, they have a common initial vertex $A_1(a_1,0) = B_0(a_1,0)$ and a common final vertex $A_{2\upsilon+1} = B_{2\nu}$. The latter condition is equivalent to assuming that

$$\sum_{j=2}^{\upsilon+1} a_j = \sum_{j=1}^{\nu} c_j \;\text{ and }\; \sum_{j=1}^{\upsilon} b_j = \sum_{j=1}^{\nu} d_j.$$

Therefore, both F_m^c and G_n^c are contained within a rectangle. They may cross each other or some of their segments may overlap and to reflect upon this, we will say that the two curves are in a general position. The ecdfs $F_m(x)$ and $G_n(x)$ and their counterparts F_m^c, G_n^c that are stepwise curves in $\mathbb{R}^2$ are illustrated in Figure 4.1.

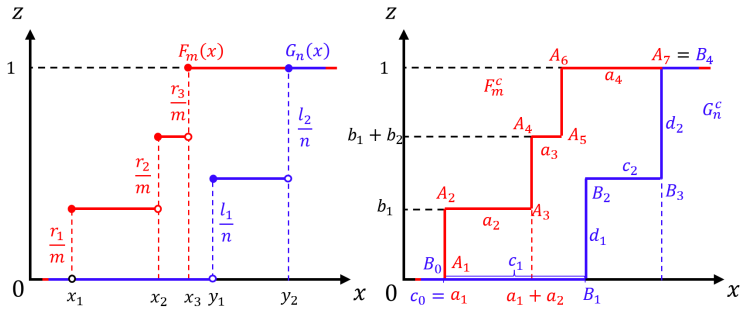


Figure 4.1: Graphs of F_m, G_n and planar curves F_m^c and G_n^c for $\upsilon = 3$ and $\nu = 2$.

Let us also note that F_m^c in Figure 4.1 divides the plane $\mathbb{R}^2$ into two open sets, the (strict) epigraph and hypograph of F_m^c denoted in the general $\mathbb{R}^{k+1}$ case $k \geq 1$, correspondingly by $\mathcal{U}_{F_m^c}$ and $\mathcal{L}_{F_m^c}$.

Given the two unknown cdfs $F(x)$ and $G(x)$, $x \in \mathbb{R}^k$ behind the random samples

$\boldsymbol{X}_m$ and $\boldsymbol{Y}_n$, in order to test the null hypothesis $H_0 : F(x) = G(x)$ for all x, against the alternative $H_1 : F(x) \neq G(x)$. for at least one x, we propose to use the Hausdorff distance $H_\rho(F_m^c, G_n^c)$ between the planar curves F_m^c and G_n^c that correspond to the ecdfs, F_m and G_n. We specify ρ to be the Chebyshev distance ρ_∞ defined in (4.2.4), i.e. we take $\rho = \rho_\infty$. The rationale behind this choice can be similarly justfied as in Section 3.2.2 and is two-fold. First, as we show in Theorem 4.5, $H_{\rho_\infty}(F_m^c, G_n^c)$ coincides with the edge of the largest hypercube that can be fitted between the planar curves F_m^c and G_n^c in $\mathbb{R}^{k+1}$ that represent staircases in $\mathbb{R}^{k+1}$. This motivates geometrically the choice $\rho = \rho_\infty$ which also leads to more efficient evaluation of $H_{\rho_\infty}(F_m^c, G_n^c)$, illustrated in Section 4.3 (see Lemmas 4.23 and 4.39 therein). Second, the choice $\rho = \rho_\infty$ allows us to establish a potentially fruitful connection between the Hausdorff and the Lévy–Prokhorov metrics given by Proposition 4.9. For brevity, we will drop the subscript ρ_∞ from $H_{\rho_\infty}(F_m^c, G_n^c)$, write $H(F_m^c, G_n^c)$ and also use the shorter notation $\mathcal{H}_{m,n}$. In what follows, we will refer to the latter Hausdorff distance as the (two-sample) Hausdorff test statistic, or simply the H test. As we noted in the introduction, the two-sample case considered here is fundamentally different from the one-sample H test, because both of the ecdfs F_m^c and G_n^c stem from unknown underlying distributions, in contrast to the one-sample case, where an explicit null cdf is hypothesized.

4.2.3 Properties of the $\mathcal{H}_{m,n}$ Statistic: the Multivariate Case

In this section, we will establish some general properties of the proposed statistic $H(F_m^c, G_n^c)$ in the multivariate case, i.e., when F_m^c and G_n^c are planar curves in $\mathbb{R}^{k+1}$, $k \geq 1$. Let us note that the sup-inf in (4.2.5) is always achievable for two compact sets, therefore the Hausdorff metric is always well-defined. However, the sets F_m^c and G_n^c are not bounded, nor is their symmetric difference $F_m^c \triangle G_n^c := (F_m^c - G_n^c) \cup (G_n^c - F_m^c)$ when $k \geq 2$, where the sup-inf in (4.2.5) can be achieved. In order to mitigate this problem, we alternatively consider two functions $\hat{F}_m$ and $\hat{G}_n$ that truncate F_m and G_n, with corresponding planar curves $\hat{F}_m^c$ and $\hat{G}_n^c$, such that their symmetric difference $\hat{F}_m^c \triangle \hat{G}_n^c$ is bounded.

Lemma 4.3. *For $M > 0, x \in \mathbb{R}^k$, let us define two functions $\hat{F}_m(x; M)$ and $\hat{G}_n(x; M)$*

truncating F_m and G_n as follows

$$\hat{F}_m(x;M) = \mathbb{1}(x \preceq M \cdot I)F_m(x) + [1 - \mathbb{1}(x \preceq M \cdot I)],$$
$$\hat{G}_n(x;M) = \mathbb{1}(x \preceq M \cdot I)G_n(x) + [1 - \mathbb{1}(x \preceq M \cdot I)],$$

where $I = (1,\ldots,1)^T \in \mathbb{R}^k$. There always exists $M_0 > 0$, such that $\forall\, M > M_0$,

$$\mathcal{H}_{m,n} = H(\hat{F}_m^c(M), \hat{G}_n^c(M)), \tag{4.2.10}$$

where $\hat{F}_m^c(M)$ and $\hat{G}_n^c(M)$ are the planar curves of $\hat{F}_m(x;M)$ and $\hat{G}_n(x;M)$ respectively.

Therefore, to compute $\mathcal{H}_{m,n}$, we can instead compute $H(\hat{F}_m^c(M), \hat{G}_n^c(M))$, for a suitably large choice of M. Since the symmetric difference $\hat{F}_m^c(M) \Delta \hat{G}_n^c(M)$ is always bounded, Lemma 4.3 provides a convenient framework for us to consider further properties of $\mathcal{H}_{m,n}$ and generalize results in Section 3.2.2. We will start with generalizing Lemmas 3.6 and 3.8.

Lemma 4.4. *The Hausdorff distance*

$$\begin{aligned}\mathcal{H}_{m,n} &= \max\left[\sup_{A\in F_m^c}\inf_{B\in G_n^c}\rho_\infty(A,B), \sup_{B\in G_n^c}\inf_{A\in F_m^c}\rho_\infty(A,B)\right] \\ &= \sup_{A\in F_m^c}\inf_{B\in G_n^c}\rho_\infty(A,B) = \sup_{B\in G_n^c}\inf_{A\in F_m^c}\rho_\infty(A,B)\end{aligned}$$

Lemma 4.4 is important as it significantly simplifies the evaluation of $\mathcal{H}_{m,n}$ requiring only one sup-inf form (instead of two). Next we formulate Theorem 4.5, which states that the Hausdorff distance with respect to the metric ρ_∞ can be expressed as the edge of the maximum hypercube that can be fitted between the curves, F_m^c and G_n^c.

Theorem 4.5. *Denote by $S(Q,d) = \{P \in \mathbb{R}^{k+1} : \rho_\infty(P,Q) \le d/2\}$ the hypercube with edge d and center at the point Q. If the hypercube $S(Q,d)$ can be inserted between the curves F_m^c and G_n^c, so that it does not overlap with the sets $\mathcal{U}_{F_m^c} \cap \mathcal{U}_{G_n^c}$ and $\mathcal{L}_{F_m^c} \cap \mathcal{L}_{G_n^c}$, then for the Hausdorff distance, we have*

$$H(F_m^c, G_n^c) \ge d. \tag{4.2.11}$$

Furthermore,

$$H(F_m^c, G_n^c) = \sup\{d : S(Q,d) \cap [(\mathcal{U}_{F_m^c} \cap \mathcal{U}_{G_n^c}) \cup (\mathcal{L}_{F_m^c} \cap \mathcal{L}_{G_n^c})] = \emptyset\} \tag{4.2.12}$$

or alternatively

$$H(F_m^c, G_n^c) = \sup\{d : S(Q,d) \subset \mathcal{G}\}, \tag{4.2.13}$$

where $\mathcal{G} = \mathbb{R}^{k+1} \setminus \{[(\mathcal{U}_{F_m^c} \cap \mathcal{U}_{G_n^c}) \cup (\mathcal{L}_{F_m^c} \cap \mathcal{L}_{G_n^c})]\}$ *is the area between two planar curves* F_m^c *and* G_n^c.

Theorem 4.5 is an important result, as it gives a geometric interpretation of the distance $H(F_m^c, G_n^c)$, which links it to the Lévy-Prokhorov metric (cf. Proposition 4.9), and also provides effcient means for computing $H(F_m^c, G_n^c)$ (cf. Lemma 4.15). In addition, based on the latter geometric interpretation, it is not difficult to see that all the observations $(X_1, \ldots, X_m)$ and $(Y_1, \ldots, Y_n)$ directly influence $H(F_m^c, G_n^c)$, in contrast to the run/rank tests (such as KS, CvM and AD), which depend solely on the relative ordering of $(X_1, \ldots, X_m)$ and $(Y_1, \ldots, Y_n)$. Furthermore, $H(F_m^c, G_n^c)$ is Lipschitz continuous as a point function of the observations, as stated in the following theorem.

Theorem 4.6. *Let* $\check{\mathbf{X}}_m = \{\check{X}_1, \ldots, \check{X}_m\}$ *and* $\check{\mathbf{Y}}_n = \{\check{Y}_1, \ldots, \check{Y}_n\}$ *be two arbitrary random samples in* $\mathbb{R}^k$ *with corresponding empirical cdfs* $\check{F}_m$ *and* $\check{G}_n$ *and planar curves* $\check{F}_m^c$ *and* $\check{G}_n^c$. *We have*

$$|H(F_m^c, G_n^c) - H(\check{F}_m^c, \check{G}_n^c)| \leq 4 \max\left\{ \max_{1 \leq i \leq m} \rho_\infty^*(X_i, \check{X}_i), \max_{1 \leq j \leq n} \rho_\infty^*(Y_j, \check{Y}_j) \right\}, \tag{4.2.14}$$

where ρ_1^* *is the metric defined by* ρ_∞ *(cf.* (4.2.4)*) restricted on* $\mathbb{R}^k$.

One consequence of Theorem 4.6 gives a key advantage of the $H(F_m^c, G_n^c)$ test over other rank/run tests, which is its robustness to small perturbations in the sample values. The latter guarantees that the goodness-of-fit testing result under $\mathcal{H}_{m,n}$ are stable when samples are noise-contaminated, which may be due to many unavoidable practical reasons, such as measurement errors, rounding errors, and privacy protection (see Avella-Medina, 2021). For example, the noise induced by rounding errors, which typically leads to ties appearing in the data, can substantially inflate the type I error when applying a rank statistic (see e.g. Schröer and Trenkler,

1995). The robustness of $\mathcal{H}_{m,n}$ with respect to small perturbations is illustrated in the following simple example.

Example 4.7. Let $\boldsymbol{X}_m$ and $\boldsymbol{Y}_n$ have realizations $\boldsymbol{X}_m = \{0, 0.5, 1\}$ and $\boldsymbol{Y}_n = \{0, 0.5, 1\}$. Let $\check{\boldsymbol{X}}_m = \{0, 0.4999, 1\}$ and $\check{\boldsymbol{Y}}_n = \{0, 0.5001, 1\}$ be noise-contaminated samples of $\boldsymbol{X}_m$ and $\boldsymbol{Y}_n$. Clearly, the values of the H and KS statistics with respect to the original samples $\boldsymbol{X}_m$ and $\boldsymbol{Y}_n$ are $\mathcal{H}_{m,n} = 0$ and $\mathcal{D}_{m,n} = 0$, whereas the values with respect to $\check{\boldsymbol{X}}_m$ and $\check{\boldsymbol{Y}}_n$ are $\check{\mathcal{H}}_{m,n} = 0.0002$ and $\check{\mathcal{D}}_{m,n} = \frac{1}{3}$. The latter values are graphically illustrated in Figure 4.2, where the cdfs of $\check{\boldsymbol{X}}_m$ and $\check{\boldsymbol{Y}}_n$ coincide except at the middle observation. As can be seen, $\check{\mathcal{H}}_{m,n} = 0.0002$ is the side of the maximum square that can be fitted between $\check{F}^c_m$ and $\check{G}^c_n$ (cf. Theorem 4.5).

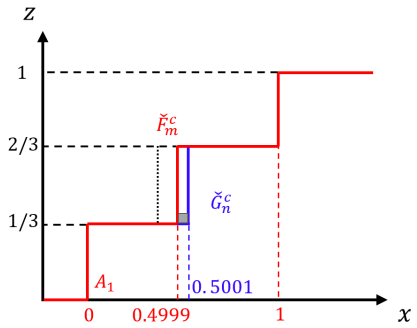


Figure 4.2: The effect of a small perturbation in the samples as in Example 4.7, on $\mathcal{H}_{m,n}$ and $\mathcal{D}_{m,n}$; $\check{\mathcal{D}}_{m,n} = 1/3$ – the dotted line, against the robust $\check{\mathcal{H}}_{m,n} = 0.0002$ - the side of the shaded rectangle which is much closer to $\mathcal{H}_{m,n} = \mathcal{D}_{m,n} = 0$.

In Theorem 4.25 of Section 4.3.2, we further establish the robustness of the p-values of $\mathcal{H}_{m,n}$, referred to as the qualitative robustness.

Next, we establish a connection between the Hausdorff and the Lévy-Prokhorov metrics. For the purpose, let us recall the definition of the Lévy-Prokhorov metric, which measures the distance between two probability measures.

Definition 4.8 (Lévy-Prokhorov metric)**.** For any subset $\mathcal{A} \subset \mathbb{R}^k$, its ε-neighbor (with respect to the metric ρ_∞) is defined as

$$\mathcal{A}^\varepsilon = \{y \in \mathbb{R}^k : \exists\, x \in \mathcal{A}, \rho_\infty(x, y) < \varepsilon\}$$

Denote by $\mathfrak{B}$ the collection of all Borel sets of $\mathbb{R}^k$, by μ_1 and μ_2 two probability measures defined on $(\mathbb{R}^k, \mathfrak{B})$. Then the Lévy-Prokhorov metric, measuring the distance

between μ_1 and μ_2, is defined as

$$\begin{aligned}\rho_{LP}(\mu_1,\mu_2) \equiv \rho_{LP}(\mu_1,\mu_2;\mathfrak{B}) = \inf\{\varepsilon>0 :\\ \mu_1(\mathcal{A}) \le \mu_2(\mathcal{A}^\varepsilon)+\varepsilon, \text{ and } \mu_2(\mathcal{A}) \le \mu_1(\mathcal{A}^\varepsilon)+\varepsilon, \forall \mathcal{A}\in\mathfrak{B}\}\end{aligned} \tag{4.2.15}$$

The following proposition expresses the Hausdorff metric in the form of a Lévy-Prokhorov metric. A Hausdorff expression of the Lévy-Prokhorov metric is given in Rachev (1981).

Proposition 4.9. *Denote by $\mu_{\boldsymbol{X}_m}$ and $\mu_{\boldsymbol{Y}_n}$ the empirical measures with respect to the samples $\boldsymbol{X}_m$ and $\boldsymbol{Y}_n$, i.e. $\mu_{\boldsymbol{X}_m} = \frac{1}{m+n}\sum_{l=1}^m \delta_{X_l}$ and $\mu_{\boldsymbol{Y}_n} = \frac{1}{m+n}\sum_{l=1}^n \delta_{Y_l}$, where δ_x is the Dirac measure, i.e. for any set $\mathcal{A}\subset\mathbb{R}^k$,*

$$\delta_x(\mathcal{A}) = \mathbb{1}_{\mathcal{A}}(x),$$

where $\mathbb{1}_{\mathcal{A}}(\cdot)$ is the indicator of the set $\mathcal{A}$. Denote by $\mathfrak{B}_i = \left\{\left\{x\in\mathbb{R}^k : x \preceq_i y\right\} : y\in\mathbb{R}^k\right\}$, then the Hausdorff statistic can also be expressed as

$$\begin{aligned}H(F^c_{m,i},G^c_{n,i}) = \rho_{LP}(\mu_{\boldsymbol{X}_m},\mu_{\boldsymbol{Y}_n};\mathfrak{B}_i) = \inf\{\varepsilon>0 : \mu_{\boldsymbol{X}_m}(\mathcal{A}) \le \mu_{\boldsymbol{Y}_n}(\mathcal{A}^\varepsilon)+\varepsilon,\\ \text{and } \mu_{\boldsymbol{Y}_n}(\mathcal{A}) \le \mu_{\boldsymbol{X}_m}(\mathcal{A}^\varepsilon)+\varepsilon, \forall\mathcal{A}\in\mathfrak{B}_i\}.\end{aligned} \tag{4.2.16}$$

Note that since each class $\mathfrak{B}_i$ for $1\le i\le 2^k$ generates the sigma algebra $\mathfrak{B}$, therefore $\mathfrak{B}_i\subset\mathfrak{B}$, and it is clear that the Hausdorff metric $H(F^c_{m,i},G^c_{n,i})$ is therefore a degenerate form of the Lévy-Prokhorov metric $\rho_{LP}(\mu_{\boldsymbol{X}_m},\mu_{\boldsymbol{Y}_n})$, which also results in Corollary 4.10.

Corollary 4.10. $H(F^c_{m,i},G^c_{n,i}) \le \rho_{LP}(\mu_{\boldsymbol{X}_m},\mu_{\boldsymbol{Y}_n})$.

The next example shows that the Hausdorff metric is a different metric from the Lévy-Prokhorov metric, in other words, the equality sign does not hold in general.

Example 4.11. Let $\boldsymbol{X}_m=\{1\}$, $\boldsymbol{Y}_n=\{0,2\}$. One easily checks that $H(F^c_m,G^c_n)=0.5$, whereas $\rho_{LP}(\mu_{\boldsymbol{X}_m},\mu_{\boldsymbol{Y}_n})\ge 1$. Indeed, take $A=\{1\}$. Then for any $\varepsilon<1$, $A^\varepsilon=[1-\varepsilon,1+\varepsilon]$ which contains neither 0 nor 2, and therefore $\mu_{\boldsymbol{Y}_n}(A^\varepsilon)=0$. Thus $\mu_{\boldsymbol{X}_m}(A)=1>\mu_{\boldsymbol{Y}_n}(A^\varepsilon)+\varepsilon=\varepsilon$ for any $\varepsilon<1$.

The Hausdorff statistic $H(F^c_{m,i},G^c_{n,i})$ depends on the component-wise order $\preceq_i$. Alternatively, we could also define the following statistic, which is independent of

the choice of $\preceq_i$, in analogy to the multivariate generalization of the KS test by Peacock (1983),

$$\begin{aligned}\mathcal{H}^{\circ}_{m,n} &= \max_{1\leq i\leq 2^k} H(F^c_{m,i}, G^c_{n,i}) = \rho_{LP}(\mu_{\boldsymbol{X}_m}, \mu_{\boldsymbol{Y}_n}; \cup\mathfrak{B}_i) \\ &= \inf\{\varepsilon > 0 : \mu_{\boldsymbol{X}_m}(\mathcal{A}) \leq \mu_{\boldsymbol{Y}_n}(\mathcal{A}^{\varepsilon}) + \varepsilon, \text{ and} \\ &\qquad \mu_{\boldsymbol{Y}_n}(\mathcal{A}) \leq \mu_{\boldsymbol{X}_m}(\mathcal{A}^{\varepsilon}) + \varepsilon, \forall \mathcal{A} \in \cup_i \mathfrak{B}_i\}.\end{aligned} \tag{4.2.17}$$

Let us note that, $\mathcal{H}^{\circ}_{m,n}$ is also a degenerate Lévy-Prokhorov distance since $\cup_i\mathfrak{B}_i$ also generates the sigma algebra $\mathfrak{B}$.

Remark 4.12. The link between the Hausdorff and Lévy-Prokhorov metrics established in Proposition 4.9 is important since it offers a bridge to applying the Hausdorff statistic $\mathcal{H}_{m,n}$ to the context of global sensitivity analysis and variable importance measurement (see Borgonovo et al., 2025b, and discussions in Section 5.2).

Remark 4.13. In practice, given a pair of data samples in high dimension, instead of directly testing for goodness-of-fit, one may first project the data into a lower dimension and then perform the goodness-of-fit testing. If the dimension reduction does not affect the scale of the data samples, it is not difficult to show that the Hausdorff distance computed with respect to the projected data (in the lower dimension) is bounded from above by $H(F^c_{m,i^*}, G^c_{n,i^*})$ with an appropriate specification of $\preceq_{i^*}$ in the original high dimension. This statement is analogus to Corollary 4.10, since the class $\mathfrak{B}_j$ generated by the component-wise order $\preceq_j$ in the lower dimension is a subset of the original class $\mathfrak{B}_{i^*}$ generated by $\preceq_{i^*}$.

Following the geometric interpretation given by Theorem 4.5, we give some results on how to compute the proposed H statistic. For the purpose, let us introduce some notation for the vertices of F^c_m and G^c_n in the case $k > 1$, generalizing the notation given in Section 4.2.2, for the case $k = 1$. Defining the vertices in the latter case is trivial, since each vertex occurs at each point of jump discontinuity of F_m and G_n. However, a direct extension to the case $k > 1$ does not work since F_m and G_n may be discontinuous not only at points but also over entire segments in $\mathbb{R}^k$. Therefore, in order to define the vertices of F_m and G_n in the case $k > 1$, we need to select the points in $\mathbb{R}^k$ at which F_m and G_n jump with respect to all its coordinates.

We refer to the latter as points of omnidirectional jumps.

Definition 4.14. Denote by $e_i = (0,\ldots,0,1,0,\ldots,0) \in \mathbb{R}^k$ $(1 \le i \le k)$, the i-th standard basis vector in $\mathbb{R}^k$, where 1 appears exactly in the i-th position. Then x is a point of omnidirectional jump of F_m if and only if, for every i,

$$\lim_{t\downarrow 0} F_m(x+te_i) \neq \lim_{t\uparrow 0} F_m(x+te_i). \tag{4.2.18}$$

When $k=1$, every jump of F_m is automatically an omnidirectional jump. Denote by $\alpha_1,\alpha_2,\ldots,\alpha_\upsilon \in \mathbb{R}^k$ and $\beta_1,\beta_2,\ldots,\beta_\nu \in \mathbb{R}^k$ all the points of omnidirectional jumps of F_m and G_n respectively. Without loss of generality, let us further assume that both samples come from continuous distributions, i.e. there are no ties in the samples $\boldsymbol{X}_m$ and $\boldsymbol{Y}_n$. In the latter case, we define the vertices $A_1,\ldots,A_{2\upsilon}$ of F_m^c as

$$A_{2i-1} = (\alpha_i, F_m(\alpha_i-))^T, A_{2i} = (\alpha_i, F_m(\alpha_i+))^T,\ i = 1,2,\ldots,\upsilon, \tag{4.2.19}$$

which are sufficient to give a computable expression of $\mathcal{H}_{m,n}$, as shown in Lemma 4.15. A more detailed description of all the vertices of G_n^c is given in Appendix 4.A.

The following lemma generalizes Proposition 3.9 and provides an explicit expression for the Hausdorff statistic $\mathcal{H}_{m,n}$ and numerical methods to compute it given in Section 4.3.1 and Appendix 4.A.

Lemma 4.15. *Let $\mathcal{L}_l$, $l = 1,2,\ldots,2\nu$ be straight lines, correspondingly passing through each of the vertices A_l of F_m defined in* (4.2.19) *and parallel to the vector OE_0,*[1] *where $E_0 = (1,\ldots,1,-1)^T \in \mathbb{R}^{k+1}$, i.e. each $\mathcal{L}_l$ is explicitly defined as*

$$\mathcal{L}_l = \{P \in \mathbb{R}^{k+1} : P - A_l = tE_0, t \in \mathbb{R}\}. \tag{4.2.20}$$

Denote by E_l, $l = 1,2,\ldots,2\upsilon$ the points of intersection of the lines, $\mathcal{L}_l$ with the planar curve G_n^c and consider the distances $\rho_\infty(A_l,E_l)$, $l = 1,2,\ldots,2\upsilon$. Then for the Hausdorff statistic, we have

$$\mathcal{H}_{m,n} = H(F_m^c, G_n^c) = \max\{\rho_\infty(A_l,E_l), l = 1,\ldots,2\upsilon\}, \tag{4.2.21}$$

[1] In this paper, the point O denotes the zero vector in $\mathbb{R}^{k+1}$ and $\mathbf{0}$ denotes the zero vector in $\mathbb{R}^k$.

under the assumption that the domain $\mathcal{G} = \mathbb{R}^{k+1} \setminus \{[(\mathcal{U}_{F_m^c} \cap \mathcal{U}_{G_n^c}) \cup (\mathcal{L}_{F_m^c} \cap \mathcal{L}_{G_n^c})]\}$ *(cf. Theorem 4.5) is bounded in* $\mathbb{R}^{k+1}$.

When $k = 1$, it can be seen that $\mathcal{G}$ in (4.2.21) is always bounded. When $k > 1$, F_m^c and G_n^c need to be truncated according to Lemma 4.3, appropriately selecting the constant M, so that $\mathcal{G}(\hat{F}_m^c(M), \hat{G}_n^c(M))$, the area between $\hat{F}_m^c(M)$ and $\hat{G}_n^c(M)$, becomes bounded. Furthermore, the vertices of each planar curve $\hat{F}_m^c(M)$ and $\hat{G}_n^c(M)$ that are defined in Lemma 4.3 are composed of two distinct parts: (i) the original vertices of F_m^c and G_n^c, (ii) additional vertices in $\hat{F}_m^c(M)$ and $\hat{G}_n^c(M)$ generated by the truncation of F_m^c and G_n^c. Therefore in order to use (4.2.21) to compute $H(F_m, G_n)$, one needs to incorporate the second part of the vertices of $\hat{F}_m^c(M)$. The choice of M and the detailed adjustment of (4.2.21) are discussed in Appendix 4.A.

Let us note that (4.2.21) can also be applied to compute $\mathcal{H}_{m,n}^{\circ}$ defined in (4.2.17). To see this, note that $F_{m,1}^c \equiv F_m^c$, $G_{m,1}^c \equiv G_n^c$ (cf. Section 4.2.1), and therefore $\mathcal{H}_{m,n} = H(F_m^c, G_n^c) \equiv H(F_{m,1}^c, G_{n,1}^c)$. To compute the remaining $H(F_{m,i}^c, G_{n,i}^c)$ $(2 \leq i \leq 2^k)$ in (4.2.17), one only needs to modify E_0 in (4.2.20) as $E_0 = (\xi_i, -1)^T \in \mathbb{R}^{k+1}$, where $\xi_i \in \{-1, 1\}^k \subset \mathbb{R}^k$, i.e. each component of ξ_i is either -1 or 1, and such that $\xi_i \succ_i \mathbf{0}$.

4.2.4 Further Properties of $\mathcal{H}_{m,n}$ in the Univariate Case

The purpose of this section is to give some further properties of the proposed statistic $\mathcal{H}_{m,n} = H(F_m^c, G_n^c)$ in the univariate case i.e., when F_m^c and G_n^c are planar curves in $\mathbb{R}^2$, i.e. when $k = 1$. In this case, its definition does not depend on the ordering $\preceq$ considered in Section 4.2.1, and it is easy to compare $\mathcal{H}_{m,n}$ with other tests such as the KS statistic. An essential difference between the latter two is that $\mathcal{H}_{m,n}$ depends on both coordinates of the planar curves F_m^c and G_n^c, while the KS statistic depends only on their vertical coordinates. As a consequence, as established in Section 4.4, $\mathcal{H}_{m,n}$ has higher power than the KS statistic. As we also note in Remark 4.22, $\mathcal{H}_{m,n}$ coincides with the Lévy metric.

The first property we introduce is a special case of Lemma 4.15, which further simplifies the computation of $\mathcal{H}_{m,n}$. For the purpose, we modify the planar curves F_m^c and G_n^c defined in $\mathbb{R}^2$ (see Fig. 4.1) so as to ensure that they only touch but not cross each other. As we show in Lemma 4.17, such a modification does not affect the value of $\mathcal{H}_{m,n}$.

Consider the subset of the vertices $B_1,\ldots,B_{2\nu}$ of G_n^c which intersects with $\mathcal{U}_{F_m^c}$, i.e. $\{B_1,\ldots,B_{2\nu}\}\cap\mathcal{U}_{F_m^c}$. Consider also the intersection $\{A_1,\ldots,A_{2v}\}\cap\mathcal{L}_{G_n^c}$. The modified planar curve $\tilde{F}_m^c$, has vertices $(\{A_1,\ldots,A_{2v}\}\setminus(\{A_1,\ldots,A_{2v}\})\cap\mathcal{L}_{G_n^c})\cup(\{B_1,\ldots,B_{2\nu}\}\cap\mathcal{U}_{F_m^c})$. Similarly, the modified planar curve $\tilde{G}_n^c$ has vertices $\{B_1,\ldots,B_{2\nu}\}\setminus(\{B_1,\ldots,B_{2\nu}\}\cap\mathcal{U}_{F_m^c})\cup(\{A_1,\ldots,A_{2v}\}\cap\mathcal{L}_{G_n^c})$. The latter sets of vertices exclude the additional vertices at the crossing points of F_m^c and G_n^c, where $\tilde{F}_m^c$ and $\tilde{G}_n^c$ touch. Therefore these additional vertices are irrelevant in the computation of $H(F_m^c,G_m^c)$.

Remark 4.16. Let us note that if we modify $\tilde{F}_m^c$ and $\tilde{G}_n^c$ applying the above modification rules, then it can be seen that $\tilde{\tilde{F}}_m^c$ and $\tilde{\tilde{G}}_n^c$ coincide respectively with $\tilde{F}_m^c$ and $\tilde{G}_n^c$.

Denote by, $\tilde{a}_1=a_1$, $\tilde{c}_0=a_1$, $\tilde{a}_{i+1}$ and $\tilde{b}_i$, $i=1,\ldots,\tilde{v}$, and by, $\tilde{c}_j$ and $\tilde{d}_j$, $j=1,\ldots,\tilde{\nu}$, the sizes of the horizontal and vertical segments of correspondingly the staircases $\tilde{F}_m^c$ and $\tilde{G}_n^c$. Their step sizes $\tilde{a}_i$, $\tilde{b}_i$, $\tilde{c}_j$ and $\tilde{d}_j$ can be expressed in terms of the step sizes a, b, c and d of F_m^c and G_n^c for any particular pair of the cdfs F_m and G_n that are in general position. This is described in Section 4.3.1. For the vertices of $\tilde{F}_m^c$ and $\tilde{G}_n^c$, we have

$$\begin{aligned}
\tilde{A}_{2l}&=\Big(\sum_{j=1}^{l}\tilde{a}_j,\sum_{j=1}^{l}\tilde{b}_j\Big)\ l=1,2,\ldots,\tilde{v};\ \tilde{A}_{2l+1}=\Big(\sum_{j=1}^{l+1}\tilde{a}_j,\sum_{j=0}^{l}\tilde{b}_j\Big)\ l=0,1,\ldots,\tilde{v},\\
\tilde{B}_{2l}&=\Big(\sum_{j=0}^{l}\tilde{c}_j,\sum_{j=0}^{l}\tilde{d}_j\Big)\ l=0,1,\ldots,\tilde{\nu};\ \tilde{B}_{2l+1}=\Big(\sum_{j=0}^{l+1}\tilde{c}_j,\sum_{j=0}^{l}\tilde{d}_j\Big)\ l=0,1,\ldots,\tilde{\nu}-1.
\end{aligned}\tag{4.2.22}$$

The following lemma states that the Hausdorff distances between the planar curves F_m^c and G_n^c and their modified versions $\tilde{F}_m^c$ and $\tilde{G}_n^c$ coincide.

Lemma 4.17. *The Hausdorff metric* $\mathcal{H}_{m,n}=H(F_m^c,G_n^c)=H(\tilde{F}_m^c,\tilde{G}_n^c)$.

Proof. The result follows since the domain $\mathcal{G}$ is not affected by the modification of (F_m^c,G_n^c) into $(\tilde{F}_m^c,\tilde{G}_n^c)$. □

The following lemma is a special of Lemma 4.15 which further simplifies the numerical evaluation of $\mathcal{H}_{m,n}$.

Lemma 4.18. *Let,* $\mathcal{L}_l$, $l=1,2,\ldots,\tilde{v}$, *be parallel straight lines, correspondingly passing through each of the vertices* $\tilde{A}_{2l}$, $l=1,2,\ldots,\tilde{v}$, *in such a way that they cross*

the horizontal axis, at an angle of $3\pi/4$. *Denote by* $\tilde{E}_{2l}$, $l = 1,2,\ldots,\tilde{v}$ *the points of intersection of the lines,* $\mathcal{L}_l$ *with the planar curve* $\tilde{G}_n^c$ *and consider the distances* $\rho_\infty(\tilde{A}_{2l},\tilde{E}_{2l})$, $l = 1,2,\ldots,\tilde{v}$. *We have*

$$\mathcal{H}_{m,n} = H(\tilde{F}_m^c,\tilde{G}_n^c) = \max\{\rho_\infty(\tilde{A}_{2l},\tilde{E}_{2l}), l = 1,\ldots,\tilde{v}\}. \tag{4.2.23}$$

We also have

$$\mathcal{H}_{m,n} = H(\tilde{F}_m^c,\tilde{G}_n^c) = \max\{\rho_\infty(\tilde{B}_{2l-1},\tilde{D}_{2l-1}), e = 1,\ldots,\tilde{\nu})\}, \tag{4.2.24}$$

where $\tilde{D}_{2l-1}$, $l = 1,\ldots,\tilde{\nu}$ *are the points of intersection of the lines* $\mathcal{L}_l'$ *with the planar curve* $\tilde{F}_n^c$, *where the lines* $\mathcal{L}_l'$ *pass through the vertices* $\tilde{B}_{2l-1}$ *and cross the horizontal axis at an angle of* $3\pi/4$.

Let us note that the maximum in (4.2.23) is taken over only the even vertices of $\tilde{F}_m^c$ in contrast to (4.2.21) in Lemma 4.15, where the corresponding maximum is taken over all vertices of F_m^c. Therefore in many practical cases, depending on the structure of the two samples $\boldsymbol{X}_m$ and $\boldsymbol{Y}_n$, i.e. of their planar curves F_m^c and G_n^c, the evaluation of $\mathcal{H}_{m,n}$ following (4.2.23) may be significantly more efficient than if one applies (4.2.21) for $k = 1$. In fact, all the even vertices of $\tilde{F}_m^c$ are all the locally concave vertices, which are also locally the farthest from $\tilde{G}_n^c$. Therefore, the maximum in (4.2.21) can be taken over the vertices which are only locally the farthest from G_n^c, which we further illustrate in Appendix 4.A.[2]

Based on Lemmas 4.18 and Theorem 4.5, we develop an efficient method for computing $H(F_m^c,G_n^c)$ and its p-values, which we present in Sections 4.3.1 and 4.3.2. Next, we state Theorem 4.19, which is important as it gives a connection between $\mathcal{H}_{m,n}$ and the two sample Kolmogorov-Smirnov statistic, $\mathcal{D}_{m,n}$, defined as

$$\mathcal{D}_{m,n} = \rho_D(F_m,G_n) := \sup_{-\infty<x<+\infty} |F_m(x) - G_n(x)|$$

Theorem 4.19. *We have*

$$H(F_m^c(\omega),G_n^c(\omega)) \le \rho_D(F_m(\omega),G_n(\omega)) \tag{4.2.25}$$

[2]A vertex $A \in G_n^c$ is locally concave iff $S(A,\delta)\cap\mathcal{L}_{G_n^c}$ is convex for some $\delta > 0$, and is locally convex iff $S(A,\delta)\cap\mathcal{U}_{G_n^c}$ is convex for some $\delta > 0$.

for every $\omega \in \Omega$. *Furthermore,*

$$H(F_m^c(\omega), G_n^c(\omega)) = \rho_D(F_m(\omega), G_n(\omega)) \tag{4.2.26}$$

for every $\omega \in \Omega$, *if*

$$\min_{2\leq i\leq \upsilon}(x_i - x_{i-1}) \geq 1 \text{ and } \min_{2\leq i\leq \nu}(y_i - x_{i-1}) \geq 1 \tag{4.2.27}$$

Remark 4.20. It can be directly seen that conditions (4.2.27) is met by the class of discrete integer valued distributions, such as Poisson, Negative Binomial, and some other distributions. In the latter case following (4.2.27), the $\mathcal{H}_{m,n}$ and $\mathcal{D}_{m,n}$ statistics coincide. However, there are important discrete distributions which do not meet conditions (4.2.27) and $\mathcal{H}_{m,n}$ and $\mathcal{D}_{m,n}$ do not coincide. It is worth noting also that when the random samples $\boldsymbol{X}_m$ and $\boldsymbol{Y}_n$ are continuous, in general, $\mathcal{H}_{m,n} \neq \mathcal{D}_{m,n}$.

As a consequence of Theorem 4.19, we can now formulate Corollary 4.21 which states that the cdf of $\mathcal{D}_{m,n}$ is dominated by the cdf of $\mathcal{H}_{m,n}$.

Corollary 4.21. *For any* $q \in [0,1]$, *we have*

$$\mathbb{P}(\mathcal{D}_{m,n} \leq q) \leq \mathbb{P}(\mathcal{H}_{m,n} \leq q).$$

Remark 4.22. Let us note that following Proposition 4.9, the univariate $\mathcal{H}_{m,n}$ statistic can be expressed as a degenerate Lévy-Prokhorov metric and therefore as a Lévy metric (cf. (3.2.14)). Therefore relations (4.2.25) and (4.2.26) can also be obtained from the property of the Lévy metric (cf. Lemma 3.46).

4.3 Evaluating $\mathcal{H}_{m,n}$ and its p-values

To test the null H_0 using the Hausdorff statistic $\mathcal{H}_{m,n} = H(F_m^c, G_n^c)$ in the general case i.e., when F_m^c and G_n^c are planar curves in $\mathbb{R}^{k+1}$ for $k \geq 1$, one needs numerical methods to efficiently evaluate it and its p-values. Such methods are presented here for the case $k = 1$ and 2.

4.3.1 Evaluating $\mathcal{H}_{m,n}$ when $k = 1$

In this section, we introduce two methods to evaluate $\mathcal{H}_{m,n}$ when $k = 1$, the projection and the transformation method. The latter is based on transforming the

coordinate system, which is easy to understand geometrically but is more difficult to generalize to higher dimensions, i.e. for $k > 1$. The projection method is slightly more complex but can be generalized to the case $k > 1$, as we demonstrate in Appendix 4.A.

The projection method for computing the value of $H(F_m^c, G_n^c)$, for a fixed pair F_m^c, and G_n^c is based on Lemmas 4.17 and 4.18. From the latter lemmas, it follows that

$$H(F_m^c, G_n^c) = H(\tilde{F}_m^c, \tilde{G}_n^c) = \max\{\rho_\infty(\tilde{A}_{2l}, \tilde{E}_{2l}), l = 1, \ldots, \tilde{\upsilon}\} \tag{4.3.1}$$

and therefore we can focus on finding $H(\tilde{F}_m^c, \tilde{G}_n^c)$ which is more convenient since, $\tilde{F}_m^c$ with vertices,

$$\{A_1, \ldots, A_{2\upsilon}\}/(\{A_1, \ldots, A_{2\upsilon}\} \cap \mathcal{L}_{G_n^c}) \cup (\{B_1, \ldots, B_{2\nu}\} \cap \mathcal{U}_{F_m^c}) \tag{4.3.2}$$

lies entirely in $\mathcal{U}_{\tilde{G}_n^c}$, i.e. $\tilde{F}_m^c \in \mathcal{U}_{G_n^c}$, and G_n^c with vertices

$$\{B_1, \ldots, B_{2\nu}\}/(\{B_1, \ldots, B_{2\nu}\} \cap \mathcal{U}_{F_m^c}) \cup (\{A_1, \ldots, A_{2\upsilon}\} \cap \mathcal{L}_{G_n^c}) \tag{4.3.3}$$

lies entirely in $\mathcal{L}_{\tilde{F}_m^c}$, i.e. $\tilde{G}_n^c \in \mathcal{L}_{\tilde{F}_m^c}$. In other words, $\tilde{F}_m^c$ lies "above" $\tilde{G}_n^c$, i.e. $\tilde{F}_m^c$ and $\tilde{G}_n^c$ may touch but not cross each other, in contrast to F_m^c and G_n^c, which are assumed in general position, i.e. may cross (see Figure 4.3). This property of $\tilde{F}_m^c$ and $\tilde{G}_n^c$ simplifies the numerical evalutation of $\mathcal{H}_{m,n}$.

Thus, following (4.3.1), in order to find all the squares, $S(Q, d)$ that fit and touch $\tilde{F}_m^c$ and $\tilde{G}_n^c$, but do not stretch into either $\mathcal{U}_{\tilde{F}_m^c}$ or $\mathcal{L}_{\tilde{G}_n^c}$, pass bisections through the vertices $\tilde{A}_{2l}$, $l = 1, \ldots, \tilde{\upsilon}$ of the 90-degree angles, formed by the staircase curve $\tilde{F}_m^c$ (See Figure 4.3), and find their points of intersection, $\tilde{E}_{2l}$, $l = 1, \ldots, \tilde{\upsilon}$ with the curve $\tilde{G}_n^c$. The point $\tilde{E}_{2l}$, with coordinates $(x_{\tilde{E}_{2l}}, z_{\tilde{E}_{2l}})$ defines the side d_l of the square $S(Q, d_l)$, i.e. $d_l = x_{\tilde{E}_{2l}} - x_{\tilde{A}_{2l}}$. Following Lemma 4.18, the Hausdorff distance $H(F_m^c, G_n^c)$ is then equal to the $\max\limits_{1 \le l \le \tilde{\upsilon}} \{d_l\}$.

The latter maximum can be found applying Lemma 4.23 which requires projecting the vertices of $\tilde{F}_m^c$ and $\tilde{G}_n^c$ onto the horizontal axis, as illustrated in Figures 4.3 and 4.4. The projection method for computing $\mathcal{H}_{m,n}$ can now be summarized as follows.

Step 1. If $\tilde{\nu} > \tilde{\upsilon}$, perform an appropriate change of the coordinate system, and

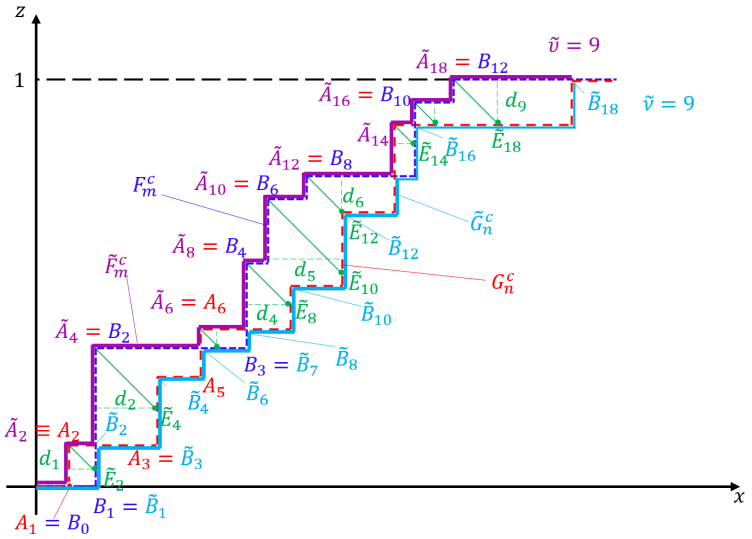


Figure 4.3: The curves $\tilde{F}_m^c$ and $\tilde{G}_n^c$, and the Hausdorff distance between them $H(\tilde{F}_m^c, \tilde{G}_n^c) = \max\limits_{1 \le l \le 9} d_l$

recompute the coordinates of all the vertices of $\tilde{F}_m^c$ and $\tilde{G}_n^c$, so that the latter change positions, i.e., $\tilde{G}_n^c$ takes position above $\tilde{F}_m^c$. If $\tilde{\nu} \le \tilde{\upsilon}$, proceed to Step 2. Therefore, Step 1 ensures that the curve with smaller number of vertices is above the curve with higher number of vertices.

Step 2. Project the vertices of $\tilde{F}_m^c$ with even indexes, i.e. the vertices $\tilde{A}_{2l}\left(\sum_{i=1}^{l}\tilde{a}_i, \sum_{i=1}^{l}\tilde{b}_i\right)^T$, $l = 1, \ldots, \tilde{\upsilon}$ onto the horizontal axis Ox, in direction of the line $x + z = 0$ in the xOz coordinate system and denote these projections with $\lambda_{\tilde{A}_{2l}}$. It is easy to see that,

$$\lambda_{\tilde{A}_{2l}} = \sum_{i=1}^{l}\left(\tilde{a}_i + \tilde{b}_i\right).$$

Step 3. Similarly as in Step 2, project all the vertices of $\tilde{G}_n^c$ and denote their projections on the abscissa Ox, by $\lambda_{\tilde{B}_l}$, $l = 1, 2, \ldots, 2\tilde{\nu}$. Clearly,

$$\lambda_{\tilde{B}_{2j}} = \sum_{k=1}^{j}\left(\tilde{c}_k + \tilde{d}_k\right), \quad \text{for} \quad j = 1, 2, \ldots, \tilde{\nu} \quad \text{and}$$

$$\lambda_{\tilde{B}_{2j+1}} = \sum_{k=1}^{j+1}\tilde{c}_k + \sum_{k=1}^{j}\tilde{d}_k, \quad \text{for} \quad j = 0, 1, 2, \ldots, \tilde{\nu} - 1.$$

Step 4. Define the semi-open intervals (open from the left and closed from the right) as:

$$\Delta_j = \left(\lambda_{\tilde{B}_{j-1}}, \lambda_{\tilde{B}_j}\right], \quad j = 1, 2, \ldots, 2\tilde{\nu}. \tag{4.3.4}$$

Order all the points from the set $\left\{\begin{array}{l}\lambda_{\tilde{B}_j}, j=0,1,2,\ldots,2\tilde{\nu}\\ \lambda_{\tilde{A}_{2l}}, l=1,2,\ldots,\tilde{v}\end{array}\right\}$ in a non-decreasing order. Steps 2-4 are illustrated graphically in Figure 4.4.

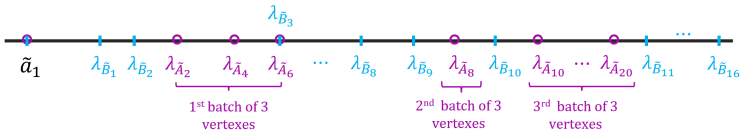


Figure 4.4: Graphical illustration of Steps 2-4. Blue tick marks indicate the points $\lambda_{\tilde{B}_i}$, $i=0,1,2,\ldots,2\tilde{\nu}$, red circles denote the points $\lambda_{\tilde{A}_{2l}}$, $l=1,2,\ldots,\tilde{v}$, and $\lambda_{\tilde{A}_1} \equiv \lambda_{\tilde{B}_0} \equiv a_1$.

Step 5. We will start the description of this step with some preliminaries. Let us consider only the intervals, Δ_i, defined as in (4.3.4), which contain at least one point from the set of projections $\{\lambda_{\tilde{A}_{2l}}, l=1,2,\ldots,\tilde{v}\}$. Order their corresponding indexes in an increasing order and denote them by $i_1 < i_2 < \ldots < i_r$, where r is the number of such intervals, $1 \le r \le 2\nu$. In the special case of Figure 4.4, $r=3$, $i_1=3$, $i_2=10$ and $i_3=11$.

Consider the i-th such interval, Δ_i, where $i \in \{i_1, i_2, \ldots, i_r\}$. We will say that the vertices, $\tilde{A}_{2s}, \tilde{A}_{2(s+1)}, \tilde{A}_{2(s+2)}, \ldots, \tilde{A}_{2t}$, with consecutive even indexes, form a batch of size $t-s+1$, if their projections, $\lambda_{\tilde{A}_{2s}}, \lambda_{\tilde{A}_{2(s+1)}}, \lambda_{\tilde{A}_{2(s+2)}}, \ldots, \lambda_{\tilde{A}_{2t}}$ on the Ox axis, belong to the interval $\Delta_i = \left(\lambda_{\tilde{B}_{i-1}}, \lambda_{\tilde{B}_i}\right]$, for some $i \in \{i_1, i_2, \ldots, i_r\}$. We will call the vertex, A_{2s}, the initial (first) vertex of the batch and A_{2t}, its final (last) vertex. We will also call the point $\lambda_{\tilde{B}_i}$, the smallest upper bound of type $\tilde{G}_n^c$, of the projections in the batch. One can easily see that, if $\tilde{A}_{2s}, \tilde{A}_{2(s+1)}, \ldots, \tilde{A}_{2t}$ is a batch with smallest upper bound, $\lambda_{\tilde{B}_i}$, then the open interval $\left(\lambda_{\tilde{A}_{2s}}, \lambda_{\tilde{A}_{2t}}\right)$ does not contain any projection from the set, $\{\lambda_{\tilde{B}_j}, j=1,2,\ldots,2\tilde{\nu}\}$. In what follows, it will be convenient to use the notation, $\bar{A}_i$, for the batch of vertices, $\tilde{A}_{2s}, \tilde{A}_{2(s+1)}, \ldots, \tilde{A}_{2t}$, whose projections $\lambda_{\tilde{A}_{2s}}, \lambda_{\tilde{A}_{2(s+1)}}, \ldots, \lambda_{\tilde{A}_{2t}}$ belong to the interval Δ_i, $i \in \{i_1, i_2, \ldots, i_r\}$.

Since $\rho_\infty(\tilde{A}_{2l}, \tilde{E}_{2l}) = \inf_{B \in \tilde{G}_n^c} \rho_\infty(\tilde{A}_{2l}, B)$, $l=1,\ldots,\tilde{v}$, to compute the maximum in (4.2.23) of Lemma 4.18, i.e. the Hausdorff distance $\mathcal{H}_{m,n} = H(\tilde{F}_m^c, \tilde{G}_n^c)$, we can compute

$$\max_{i \in \{i_1,\ldots,i_r\}} \max_{A \in \bar{A}_i} \inf_{B \in \tilde{G}_n^c} \rho_\infty(A,B). \tag{4.3.5}$$

The following lemma gives an explicit expression for the second maximum in (4.3.5) and therefore for $\mathcal{H}_{m,n}$.

Lemma 4.23. *For the distance,* $\max\limits_{A\in\bar{A}_i}\inf\limits_{B\in\tilde{G}_n^c}\rho_\infty(A,B)$, *we have*

$$\begin{aligned}\max_{A\in\bar{A}_i}\inf_{B\in\tilde{G}_n^c}\rho_\infty(A,B)&=\max_{s\le j\le t}\inf_{B\in\tilde{G}_n^c}\rho_\infty(\tilde{A}_{2j},B)\\&=\begin{cases}\sum\limits_{k=1}^{t}\tilde{b}_k-\sum\limits_{k=1}^{p}\tilde{d}_k=\rho_\infty\left(\tilde{A}_{2t},\tilde{E}_{2t}\right), & \text{if } i \text{ is odd, i.e. } i=2p+1\\ \sum\limits_{k=1}^{p+1}\tilde{c}_k-\sum\limits_{k=2}^{s}\tilde{a}_k=\rho_\infty\left(\tilde{A}_{2s},\tilde{E}_{2s}\right), & \text{if } i \text{ is even, i.e. } i=2(p+1)\end{cases}\end{aligned}\tag{4.3.6}$$

for some fixed $i\in\{i_1,i_2,\ldots,i_r\}$, $m\le n$ *and* p *obtained from* $i=2p+1$ *or* $i=2(p+1)$.

The Hausdorff distance, $\mathcal{H}_{m,n}$ is obtained by taking the maximum in (4.3.5) over all the r batches, $\bar{A}_{i_l}$, $l=1,2,\ldots,r$. This is implemented sequentially applying the following algorithm.

Denote by s_q and t_q, the indexes of, correspondingly the initial and final vertices, $\tilde{A}_{2s_q}$ and $\tilde{A}_{2t_q}$ of the q-th batch, $\bar{A}_{i_q}$ $q=1,2,\ldots,r$. Set, $q=1$ and find s_1, i_1 and t_1. Clearly, $\tilde{A}_2$ is the initial vertex of the first batch, i.e. $\tilde{A}_2=\tilde{A}_{2\times1}$, therefore $s_1=1$. Next, we find the index $i=i_1$, of the smallest upper bound $\lambda_{\tilde{B}_i}$, $i=1,2,\ldots,2\tilde{\nu}$, of the first batch $\bar{A}_1$. The index i_1 is obtained by sequentially checking the inequalities

$$\begin{aligned}&\lambda_{\tilde{A}_2}-\lambda_{\tilde{B}_1}=\lambda_{\tilde{A}_{2s_1}}-\lambda_{\tilde{B}_1}>0,\quad \lambda_{\tilde{A}_{2s_1}}-\lambda_{\tilde{B}_2}>0,\\&\ldots,\lambda_{\tilde{A}_{2s_1}}-\lambda_{\tilde{B}_{i_1-1}}>0,\quad \lambda_{\tilde{A}_{2s_1}}-\lambda_{\tilde{B}_{i_1}}\le0,\end{aligned}$$

starting with the index $i=1$, until the index $i=i_1$, for which $\lambda_{\tilde{B}_{i_1}}$, becomes equal to or exceeds $\lambda_{\tilde{A}_{2s_1}}$, i.e., $\lambda_{\tilde{A}_{2s_1}}\le\lambda_{\tilde{B}_{i_1}}$.

The index t_1, is then obtained in a similar way, by sequentially checking the inequalities

$$\lambda_{\tilde{B}_{i_1}}-\lambda_{\tilde{A}_2}\ge0,\quad \lambda_{\tilde{B}_{i_1}}-\lambda_{\tilde{A}_4}\ge0,\quad \ldots,\lambda_{\tilde{B}_{i_1}}-\lambda_{\tilde{A}_{2t_1}}\ge0,\quad \lambda_{\tilde{B}_{i_1}}-\lambda_{\tilde{A}_{2(t_1+1)}}<0.$$

Applying Lemma 4.23, substitute the indexes, $s=s_1$, $i=i_1$ and $t=t_1$, in (4.3.6) and compute the distance, $\max\limits_{A\in\bar{A}_1}\inf\limits_{B\in\tilde{G}_n^c}\rho_\infty(A,B)$, for the first batch. Set, $H\left(\tilde{F}_m^c,\tilde{G}_n^c\right)=\max\limits_{A\in\bar{A}_1}\inf\limits_{B\in\tilde{G}_n^c}\rho_\infty(A,B)$.

For the second batch, $\bar{A}_2$, we can directly see that its first vertex is $A_{2m_1}=$

$A_{2(n_1+1)}$. Therefore, the index i_2 is found by sequentially verifying the inequalities

$$\lambda_{\tilde{A}_{2s_2}} - \lambda_{\tilde{B}_{i_1+1}} > 0, \quad \lambda_{\tilde{A}_{2m_2}} - \lambda_{\tilde{B}_{i_1+2}} > 0, \ldots,$$
$$\lambda_{\tilde{A}_{2s_2}} - \lambda_{\tilde{B}_{i_2-1}} > 0, \quad \lambda_{\tilde{A}_{2s_2}} - \lambda_{\tilde{B}_{i_2}} \leq 0,$$

starting with the index $i = i_1 + 1$, until the index $i = i_2$, for which $\lambda_{\tilde{B}_{i_2}}$, becomes equal to or exceeds $\lambda_{\tilde{A}_{2s_2}}$, i.e., $\lambda_{\tilde{A}_{2s_2}} \leq \lambda_{\tilde{B}_{i_2}}$.

The index n_2 is found, similarly as for the first batch, from the inequalities

$$\lambda_{\tilde{B}_{i_2}} - \lambda_{\tilde{A}_{2s_2}} \geq 0, \quad \lambda_{\tilde{B}_{i_2}} - \lambda_{\tilde{A}_{2(s_2+1)}} \geq 0,$$
$$\ldots, \lambda_{\tilde{B}_{i_2}} - \lambda_{\tilde{A}_{2(t_2-1)}} \geq 0, \quad \lambda_{\tilde{B}_{i_2}} - \lambda_{\tilde{A}_{2t_2}} < 0.$$

Substitute the indexes, $s = s_2$, $i = i_2$ and $t = t_2$, in (4.3.6) and compute the distance, $\max\limits_{A\in\bar{A}_2} \inf\limits_{B\in\tilde{G}_n^c} \rho_\infty(A,B)$, for the second batch. If $H\left(\tilde{F}_m^c, \tilde{G}_n^c\right) < \max\limits_{A\in\bar{A}_2} \inf\limits_{B\in\tilde{G}_n^c} \rho_\infty(A,B)$ then set

$$H\left(\tilde{F}_m^c, \tilde{G}_n^c\right) = \max_{A\in\bar{A}_2} \inf_{B\in\tilde{G}_n^c} \rho_\infty(A,B),$$

else set $q = q+1$ and following the algorithm for the previous batches, compute the distance $\max\limits_{A\in\bar{A}_q} \inf\limits_{B\in\tilde{G}_n^c} H(A,B)$, for the q-th batch. Continue this process until reaching the last, r-th batch $\bar{A}_r$, for which $\tilde{A}_{2t_r} = \tilde{A}_{2\tilde{v}}$, i.e., $t_r = \tilde{v}$. Clearly, the required Hausdorff distance between the curves $\tilde{F}_m^c$ and $\tilde{G}_n^c$ is obtained in $H\left(\tilde{F}_m^c, \tilde{G}_n^c\right)$. This completes the description of the projection method.

Under the transformation method, the curves F_m^c and G_n^c that are assumed in general position are transformed into a new coordinate system rotating them by $\frac{\pi}{4}$ as shown in Figure 4.5. The Hausdorff distance is then computed with respect to the rotated F_m^c and G_n^c. In what follows, we will briefly introduce the idea behind this transformation method.

Let us define $T_r : \mathbb{R}^2 \to \mathbb{R}^2$ the rotation operator, such that for any $(x,z) \in \mathbb{R}^2$,

$$\begin{bmatrix} x' \\ z' \end{bmatrix} = T_r(x,y) = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x \\ z \end{bmatrix}, \tag{4.3.7}$$

i.e. it rotate the xOz plane by $\frac{\pi}{4}$ to the $x'Oz'$ plane. Based on Theorem 4.5, the Hausdorff distance is the side of the largest square fitted between F_m^c and G_n^c,

which has the largest diagonal. Therefore, one can compute $H(F_m^c, G_n^c)$ by looking at the vertical largest difference of the rotated curves $T_r(F_m^c)$ and $T_r(G_n^c)$, which corresponds to the largest diagonal.

Applying the operator T_r, the curves F_m^c and G_n^c are transformed into the functions $h_1(x')$ and $h_2(x')$, which are piecewise linear with each segment having slope ± 1 and corresponding vertices $T_r(A_i)$, $i = 1, 2, \ldots, 2\upsilon$ and $T_r(B_j)$, $j = 1, 2, \ldots, 2\nu$. Denote by $z' = h_1(x')$ and $z' = h_2(x')$ the parametric equations of the rotated curves $T_r(F_m^c)$ and $T_r(G_n^c)$, i.e. $\{(x', h_1(x')) : x' \in \mathbb{R}\} = T_r(F_m^c)$ and $\{(x', h_2(x')) : x \in \mathbb{R}\} = T_r(G_n^c)$. Then the Hausdorff distance $H(F_m^c, G_n^c)$ is obtained as

$$H(F_m^c, G_n^c) = \sup_{-\infty < x' < +\infty} \frac{1}{2} |h_1(x') - h_2(x')|. \tag{4.3.8}$$

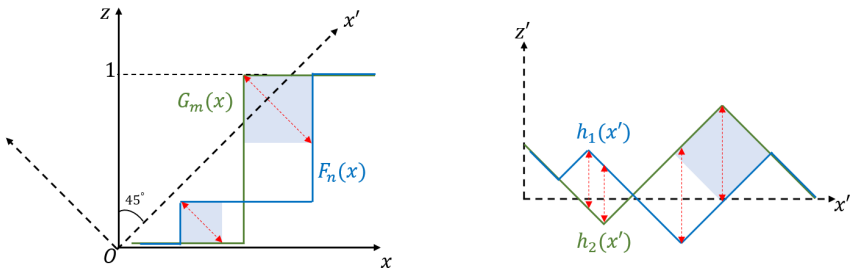


Figure 4.5: Computing $H(F_m^c, G_n^c)$ applying the transformation method.

Following Lemma 4.15, it is not difficult to show that the supremum in (4.3.8) is achieved over all x_i, $i = 1, 2, \ldots, 2\upsilon$, where x_i is the abscissa with respect to $x'Oz'$ of the transformed vertex $T_r(A_i)$. Then to compute (4.3.8), one only needs to compute $h_1(x_i) - h_2(x_i)$ sequentially, where $h_1(x_i)$ and $h_2(x_i)$ are known by definition.

We have implemented both the projection and transformation methods in C++, and their numerical efficiency is compared and discussed in Appendix 4.D. The projection method is slightly more accurate, usually providing 16 correct digits but is slower compared to the transformation method, which provides 14 correct digits and is significantly faster, as shown in Table 4.2.

The projection method outlined above for the case $k = 1$ is generalized for the case $k = 2$ in Appendix 4.A.

4.3.2 The p-values of $\mathcal{H}_{m,n}$ and its Permutation Version

The distribution of $\mathcal{H}_{m,n}$ is not easy to obtain since it depends on the underlying, (unknown) distributions F and G. This hinders the computation of its p-value

$\mathbb{P}(\mathcal{H}_{m,n} > q)$, $q \in (0,1)$, therefore its application in testing the null hypothesis H_0. In order to mitigate this difficulty, in (4.3.10) we introduce a permutation version of the $\mathcal{H}_{m,n}$ test and show in Theorems 4.24 and 4.25 that its p-value conditional on $\boldsymbol{Z}_{m+n}$ approximates asymptotically the p-value of $\mathcal{H}_{m,n}$. The p-values of the permutation H test are easy to compute by applying Proposition 4.27 (cf. (4.3.14)).

Following Section 3.8 in van der Vaart and Wellner (2023), let the random vector $\boldsymbol{R}_{m+n} = (R_1, R_2, \ldots, R_{m+n})$ be independent from the samples $\boldsymbol{X}_m$ and $\boldsymbol{Y}_n$[3] and uniformly distributed on the set of permutations of $\{1, 2, \ldots, m+n\}$, i.e. every possible realization of $\boldsymbol{R}_{m+n}$ has probability $1/(m+n)!$. Given a realization of $\boldsymbol{Z}_{m+n}$, define the permuted samples $\boldsymbol{X}_m^\dagger = \{Z_{R_1}, \ldots, Z_{R_m}\}$ and $\boldsymbol{Y}_n^\dagger = \{Z_{R_{m+1}}, \ldots, Z_{R_{m+n}}\}$ with empirical cdfs

$$F_m^\dagger(x) = \frac{1}{m}\sum_{i=1}^{m} \mathbb{1}(Z_{R_i} \preceq x),\ G_n^\dagger(x) = \frac{1}{n}\sum_{i=m+1}^{m+n} \mathbb{1}(Z_{R_i} \preceq x). \tag{4.3.9}$$

Then the corresponding permutation Hausdorff test, which we denote as $\mathcal{H}_{m,n}^\dagger$, is defined as

$$\mathcal{H}_{m,n}^\dagger = H(F_m^{\dagger c}, G_n^{\dagger c}), \tag{4.3.10}$$

where $F_m^{\dagger c}, G_n^{\dagger c}$ are the corresponding planar curves of $F_m^\dagger$ and $G_n^\dagger$, with vertices $A_{j,\cdot}^\dagger$ $(j = 1, \ldots, \upsilon^\dagger)$ and $B_{j,\cdot}^\dagger$ $(j = 1, \ldots, \nu^\dagger)$. Given the pooled sample, the p-value of the permutation statistic $\mathcal{H}_{m,n}^\dagger$ is expressed as $1 - \mathbb{P}(\mathcal{H}_{m,n}^\dagger \le q | \boldsymbol{Z}_{m+n}), q \in [0,1]$.

The essential reason to introduce the permutation statistic is the relationship between the critical levels $q_{m,n}^*(p)$ and $q_{m,n}(p)$ of the $\sqrt{\frac{mn}{m+n}}\mathcal{H}_{m,n}^\dagger$ and $\sqrt{\frac{mn}{m+n}}\mathcal{H}_{m,n}$, which are defined as

$$q_{m,n}^\dagger(p) = \inf\left\{q : \mathbb{P}\left(\sqrt{\frac{mn}{m+n}}\mathcal{H}_{m,n}^\dagger \le q | \boldsymbol{Z}_{m+n}\right) \ge p\right\},$$
$$q_{m,n}^*(p) = \inf\left\{q : \mathbb{P}\left(\sqrt{\frac{mn}{m+n}}\mathcal{H}_{m,n} \le q\right) \ge p\right\},$$

where the normalization $\sqrt{\frac{mn}{m+n}}$ guarantees that $q_{m,n}^*(p)$ is nondegenerate. If we can show that under the null, the difference between the critical levels $q_{m,n}^*(p)$ and

[3] Sometimes, $\boldsymbol{R}_{m+n}$ is considered to be defined on another probability space $(\Omega^*, \mathcal{F}^*, \mathbb{P}^*)$ and the independence is considered as on the product space of $(\Omega, \mathcal{F}, \mathbb{P})$ and $(\Omega^*, \mathcal{F}^*, \mathbb{P}^*)$. The permutation statistic is measured by $\mathbb{P}^*$, which is also called outer probability in the literature.

$q_{m,n}(p)$ is asymptotically negligible, i.e.

$$q^{\dagger}_{m,n}(p) - q^{*}_{m,n}(p) \xrightarrow{\mathbb{P}} 0, \text{ as } n, m \to \infty, \tag{4.3.11}$$

where $\xrightarrow{\mathbb{P}}$ stands for convergence in probability, we should have that the difference between the p-values of $\mathcal{H}_{m,n}$ and $\mathcal{H}^{\dagger}_{m,n}$ is also asymptotically negligible. In fact, this will be the case under the null, since by applying Theorem 15.2.3 of Lehmann and Romano (2005), one can show that under the null (4.3.11) holds universally regardless of the choice of null and the dimension k.

However, this conclusion does not imply that the $\mathcal{H}_{m,n}$ and $\mathcal{H}^{\dagger}_{m,n}$ tests lead to the same statistical inference when sample sizes are sufficiently large, since it is also often the case that the null does not hold. For this reason, in Theorems 4.24 and 4.25, we present some asymptotic results under fixed or contiguous alternative, assuming $\mathcal{H}_{m,n}$ is univariate, i.e. when $k = 1$.

Theorem 4.24. *For a fixed significance level p, let $q^{\dagger}_{m,n}(p)$ be the critical level of the statistic $\sqrt{\frac{mn}{m+n}}\mathcal{H}^{\dagger}_{m,n}$, assuming that $\boldsymbol{X}_m$ and $\boldsymbol{Y}_n$ come from the univariate distributions $F(x)$ and $G(x)$ respectively. Let $q^{*}_{m,n}(p)$ be the critical level of the statistic $\sqrt{\frac{mn}{m+n}}\mathcal{H}_{m,n}$ when the samples $\boldsymbol{X}_m$ and $\boldsymbol{Y}_n$ come from one and the same distribution $\eta F(x) + (1-\eta)G(x)$. If $F(x)$ and $G(x)$ have bounded densities, i.e. $F, G \in C^1(\mathbb{R})$ and $\frac{n}{n+m} \to \eta$, then (4.3.11) holds and the difference between the p-values of $\mathcal{H}_{m,n}$ and $\mathcal{H}^{\dagger}_{m,n}$ is also asymptotically negligible.*

Theorem 4.25. *Let $q^{\dagger}_{m,n}(p)$ and $q^{*}_{m,n}(p)$ be defined as in Theorem 4.24. Define the metric $\rho_{C^1}(F,G) = \sup_x |F(x) - G(x)| + \sup_x |\frac{d}{dx}F(x) - \frac{d}{dx}G(x)|$. Under a contiguous alternative $G(x;n) \in C^1(\mathbb{R})$, i.e. if $\rho_{C^1}(F, G(\cdot;n)) = o(1)$, then (4.3.11) holds and the difference between the p-values of $\mathcal{H}_{m,n}$ and $\mathcal{H}^{\dagger}_{m,n}$ is also asymptotically negligible.*

Theorems 4.24 and 4.25 show that the difference between $q^{\dagger}_{m,n}$ and $q^{*}_{m,n}$ is asymptotically negligible under the alternative. It is worth noting that the relation established in Theorem 4.25 describes the asymptotic behavior of the p-value of $\mathcal{H}_{m,n}$, and shows that the p-value of $\mathcal{H}_{m,n}$ does not change significantly under mild perturbation of the null. This is an important property, which is also often referred to as the qualitative robustness of a statistic (see Rieder 1982 and more recently Liu and Briol 2024). Additionally, Theorems 4.24 and 4.25 suggest that the type I

error and the power of the $\mathcal{H}_{m,n}$ and $\mathcal{H}^{\dagger}_{m,n}$ tests are asymptotically identical, as we formally show in the next theorem.

Theorem 4.26. *Assuming that F and G are defined in $\mathbb{R}^k$ ($k \geq 1$), under the null $F = G$, both the $\mathcal{H}_{m,n}$ test and its permutation version $\mathcal{H}^{\dagger}_{m,n}$ control the type I error, i.e.*

$$\mathbb{P}\left(\sqrt{\frac{mn}{m+n}}\mathcal{H}_{m,n} > q^{\dagger}_{m,n}(p)\right) \leq p \quad \textit{and} \quad \mathbb{P}\left(\sqrt{\frac{mn}{m+n}}\mathcal{H}_{m,n} > q^{*}_{m,n}(p)\right) \leq p. \tag{4.3.12}$$

Furthermore, when $k = 1$, under a fixed alternative $G \neq F \in C^1(\mathbb{R})$ or a contiguous alternative $G(x;n) \in C^1(\mathbb{R})$, i.e. $\rho_{C^1}(F, G(\cdot;n)) = O(n^{-1/2})$, for the power of $\mathcal{H}_{m,n}$ we have

$$\mathbb{P}\left(\sqrt{\frac{mn}{m+n}}\mathcal{H}_{m,n} > q^{\dagger}_{m,n}(p)\right) = \mathbb{P}\left(\sqrt{\frac{mn}{m+n}}\mathcal{H}_{m,n} > q^{*}_{m,n}(p)\right) + o(1), \tag{4.3.13}$$

where the metric $\rho_{C^1}(F,G) = \sup_x |F(x) - G(x)| + \sup_x |\frac{d}{dx}F(x) - \frac{d}{dx}G(x)|$ and $q^{*}_{m,n}$ is the critical value when both $\boldsymbol{X}_m$ and $\boldsymbol{Y}_n$ come from the null F.

Note that we have defined the vector $\boldsymbol{R}_{m+n}$ over all possible permutations of $\{1, \ldots, m+n\}$ and any further permutation among the first m or the last n elements of $\boldsymbol{R}_{m+n}$ does not change $F^{\dagger}_m$ and $G^{\dagger}_n$. Therefore, any permutation can also be equivalently considered as a split of the pooled sample $\boldsymbol{Z}_{m+n}$ into $\boldsymbol{X}^{\dagger}_m$ and $\boldsymbol{Y}^{\dagger}_n$. Denote by $\Pi(m+n)$ the set of all $C = \binom{m+n}{m}$ possible splits of $\boldsymbol{Z}_{m+n}$. One can alternatively consider $\boldsymbol{R}_{m+n}$ uniformly distributed on $\Pi(m+n)$ with equal proabaility $1/C$. Therefore, we have

Proposition 4.27. *For the permutation H test, we have*

$$\mathbb{P}(\mathcal{H}^{\dagger}_{m,n} > q | \boldsymbol{Z}_{m+n}) = \frac{1}{C}\sum_{i=1}^{C} \mathbb{1}\{\mathcal{H}^{\dagger}_{m,n} > q | \boldsymbol{R}_{m+n} = \pi_i(m+n), \boldsymbol{Z}_{m+n}\}, \tag{4.3.14}$$

where each $\pi_i(m+n) \in \Pi(m+n)$, $i = 1, \ldots, C$ represents a realization of $\boldsymbol{R}_{m+n}$ corresponding to a unique split of $\boldsymbol{Z}_{m+n}$.

Let us note that (4.3.14) is exact and computable, since given the permutation, the event $\{\mathcal{H}^{\dagger}_{m,n} \leq q\}$ in (4.3.14) can be expressed as the event of a trajectory $G^{\dagger c}_n$ (or $F^{\dagger c}_m$) non-exiting two boundaries and can be explicitly evaluated. For more details,

we refer to Appendix 4.B.

When n and m are small, all the possible realizations C, of $\boldsymbol{R}_{m+n}$ can be generated at an affordable computational cost, which leads to the exact evaluation of (4.3.14). For large n and m, the number $C = \binom{m+n}{m}$ is very large and exact evaluation of (4.3.14) becomes prohibitively time consuming. To alleviate this, we propose to use the Monte Carlo sampling of $\boldsymbol{R}_{m+n}$ to estimate the p-value, which is implemented in the numerical studies in Section 4.4.

Alternatively, the p-value of $\mathcal{H}_{m,n}$ in the univariate case $k = 1$ for large sample sizes m and n can be calculated based on the following asymptotic theorem.

Theorem 4.28. *Let $\frac{m}{m+n} \to \eta \in (0,1)$ as $m \to \infty$. Assume that $\boldsymbol{X}_m$ and $\boldsymbol{Y}_n$ come from the univariate distributions F and G respectively with bounded densities. Given the realization of Z_{m+n}, when $\boldsymbol{R}_{m+n}$ is uniformly distributed over the permutations of $\{1,2,\ldots,m+n\}$, we have*

$$\begin{aligned}\lim_{m\to\infty} \mathbb{P}\left(\sqrt{\frac{mn}{m+n}}\mathcal{H}^{\dagger}_{m,n} > x \middle| \boldsymbol{Z}_{m+n}\right) \\ = 1 - \mathbb{P}\left(|\mathbb{B}_0(t)| \le x\left(1 + e(E^{-1}(t))\right), \forall\, 0 \le t \le 1\right),\end{aligned} \tag{4.3.15}$$

where $E(x) = \eta F(x) + (1-\eta)G(x)$, with density e and inverse E^{-1} and $\mathbb{B}_0(t)$ $t \in [0,1]$ is a standard Brownian Bridge, i.e. a Gaussian process with $\mathbb{B}_0(0) = \mathbb{B}_0(1) = 0$, $\mathbb{E}[\mathbb{B}_0(t)] = 0$, $\mathbb{E}[\mathbb{B}_0(t)\mathbb{B}_0(s)] = s(1-t)$ for $0 < s < t < 1$.

In practice, the null distributions F and G are unknown. However, we could still compute the asymptotic p-value in (4.3.15), since the $e(E^{-1}(t))$ is also the reciprocal of the quantile density function (i.e. the derivative of E^{-1}), which can be estimated from the realization of $\boldsymbol{Z}_{m+n}$. The latter estimation problem is also well studied, see e.g. Soni et al. (2012) and Chesneau et al. (2016), to name only a few of the references. The estimator $\hat{\delta}(\hat{E}(t))$ can be directly substituted in (4.3.15) to obtain the asymptotic p-value.

4.4 On the Scale Dependence of $\mathcal{H}_{m,n}$

In this section, we investigate the statistical power of the two-sample H test $\mathcal{H}_{m,n} = H(F^c_m, G^c_n)$ in the univariate and bivariate cases, i.e., when F^c_m and G^c_n are planar curves in $\mathbb{R}^{k+1}$, $k = 1,2$. While $\mathcal{H}_{m,n}$ is clearly location invariant, as shown in

Chapter 3 for the one-sample case and in Example 4.29 for the two-sample case, $\mathcal{H}_{m,n}$ and its power are not invariant under scale transformation (see Figure 4.6). In this section, we explore how change of scale affects the power of the H test, and further propose rules to adjust the scale so as to optimize the power.

We formally describe the scale dependence of the statistic $\mathcal{H}_{m,n}$ as follows. For a given pair of samples, $\{X_1,\ldots,X_m\}$ and $\{Y_1,\ldots,Y_n\}$ and a scaling constant $\sigma>0$, we apply $\mathcal{H}_{m,n}$ to test the null hypothesis H_0 based on the scaled samples $\sigma\boldsymbol{X}_m=\{\sigma X_1,\ldots,\sigma X_m\}$ and $\sigma\boldsymbol{Y}_n=\{\sigma Y_1,\ldots,\sigma Y_n\}$. As illustrated in the following example, the result of testing the null H_0 will depend on σ.

Example 4.29. Let $\boldsymbol{X}_m$ and $\boldsymbol{Y}_n$ be samples from two different distributions. We use the KS, CVM, AD, W and the proposed $\mathcal{H}_{m,n}$ statistic to test the null H_0, i.e. whether the scaled samples $\sigma\boldsymbol{X}_m$ and $\sigma\boldsymbol{Y}_n$ come from the same distributions. To do so, for a fixed $\sigma>0$, we compute the p-values for all these statistics applying the permutation approach as in Proposition 4.27 using 2000 simulated permutations. We then compare the latter p-values with the significance level $p=0.05$. The power of each statistic is estimated as the frequency of rejecting the null H_0, i.e. the frequency of its p-value being less than $p=0.05$. We use the R package **twosamples** (Dowd, 2023) to evaluate the two-sample W statistic.

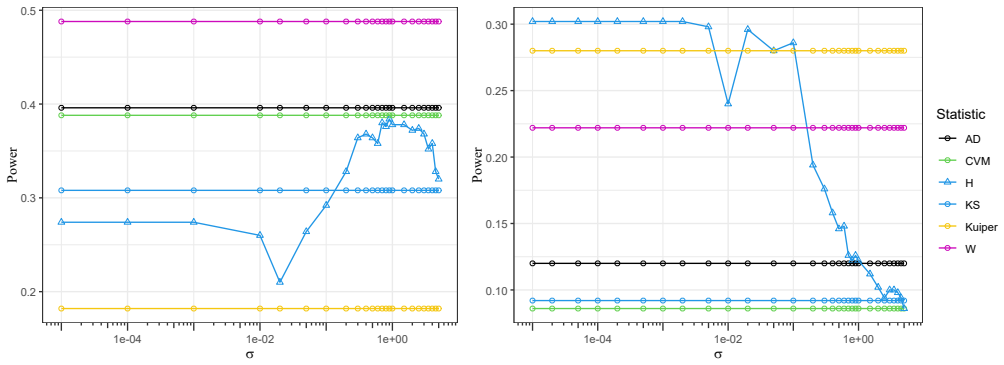


Figure 4.6: The power of KS, CvM, AD, W and $\mathcal{H}_{m,n}$ as a function of the scaling coefficient σ when $m=n=50$ and $\boldsymbol{X}_m$ and $\boldsymbol{Y}_n$ come from Exp(2) and Exp(3) (Left Panel), $N(0,2)$ and $N(0,4)$ (Right Panel).

The power of each statistic as a function of σ varied over the range $[10^{-5},5]$ is illustrated in Figure 4.6 for the case $m=n=50$ and when $\boldsymbol{X}_m$ and $\boldsymbol{Y}_n$ come from different Exponential and different Normal distributions. As can be seen, only the power of $\mathcal{H}_{m,n}$ varies with σ, whereas the powers of KS, CvM, AD and W remain

constant.

The scale dependence of the power of $\mathcal{H}_{m,n}$ is a consequence of the scale dependence of its p-value. However, the latter scale dependence is a considerable obstacle in goodness-of-fit testing with $\mathcal{H}_{m,n}$ since its results will depend on the units used to express the sample observations. To mitigate this problem which also arises for the goodness-of-fit test statistic based on the Lévy metric (see Remark 4.22 in Section 4.2.4 for its equivalence of H), Alexander (1974) proposed to first apply $E_{m+n}(\cdot)$ to each observation and then use the Lévy meausure based statistic. This means to test whether the samples $E_{m+n}(\boldsymbol{X}_m) = \{E_{m+n}(X_1), \ldots, E_{m+n}(X_m)\}$ and $E_{m+n}(\boldsymbol{Y}_n) = \{E_{m+n}(Y_1), \ldots, E_{m+n}(Y_m)\}$ come from the same distribution. Note that the observations in $E_{m+n}(\boldsymbol{X}_m)$ and $E_{m+n}(\boldsymbol{Y}_n)$ only take values $\frac{i}{m+n}$, $i = 1, \ldots, m+n$. Thus, this test now becomes a rank test, which does not depend on the scale. While this approach seems to eliminate the scale dependence, it is less appealing since it does not take into account the information contained in each observation, as shown in Theorem 4.6. Such rank transformation violates the Lipschitz continuity in Theorem 4.6 and therefore leads to loss of robustness of the statistic and a decrease in its power, which approaches that of the KS test.

A different approach to the lack of scale invariance of the H test has been proposed in Chapter 3 for the one-sample case. By viewing the H test as a function of a scale parameter σ, we have proposed a rule (3.3.11) to select σ so that the power of H is optimized. This rule explicitly uses the form of the hypothesized null distribution. We generalize this approach to the two-sample case, which is not straightforward, since in this case no information is available about the underlying null distribution.

Here we propose rules (4.4.6) and (4.4.9) to select σ, so that the H test preserves the Lipschitz continuity property of Theorem 4.6 and therefore its robustness and the scale dependence is eliminated. Furthermore, since the power of $\mathcal{H}_{m,n}$ changes with σ as we demonstrate in Example 4.29, σ should be selected so as to optimize the latter power, similarly as in the one sample case. This motivates the interpretation of σ as a scale parameter.

Let $\sigma = (\sigma^{(1)}, \ldots, \sigma^{(k)})^T \in \mathbb{R}^k$ be the scale parameter with non-negative components, i.e. $\sigma^{(i)} > 0$ for $i = 1, \ldots, k$. For any vector $x = (x^{(1)}, \ldots, x^{(k)})^T \in \mathbb{R}^k$, the

component-wise scaled vector is defined as

$$\sigma x = (\sigma^{(1)}x^{(1)}, \ldots, \sigma^{(k)}x^{(k)})^T.$$

Given samples $\boldsymbol{X}_m$ and $\boldsymbol{Y}_m$ their scaled versions are denoted by $\sigma\boldsymbol{X}_m = \{\sigma X_1, \ldots, \sigma X_m\}$ and $\sigma\boldsymbol{Y}_n = \{\sigma Y_1, \ldots, \sigma Y_n\}$. The empirical cumulative distribution functions (ecdfs) of these scaled samples, i.e. the scaled versions of F_m and G_n defined in (4.2.2) are denoted by $F_{m,\sigma}(\cdot)$ and $G_{n,\sigma}(\cdot)$. The H statistic dependent on a scale parameter σ is denoted by $\mathcal{H}_{m,n}(\sigma) = H(F^c_{m,\sigma}, G^c_{n,\sigma})$, where $F^c_{m,\sigma}$ and $G^c_{n,\sigma}$ are the planar curves of $F_{m,\sigma}(\cdot)$ and $G_{n,\sigma}(\cdot)$. We also use $\mathcal{H}^\circ_{m,n}(\sigma)$ and $\mathcal{H}^\dagger_{m,n}(\sigma)$ to denote the scaled versions of $\mathcal{H}^\circ_{m,n}$ and $\mathcal{H}^\dagger_{m,n}$ defined in (4.2.17) and (4.3.10) and view $\mathcal{H}_{m,n}(\sigma)$, $\mathcal{H}^\circ_{m,n}(\sigma)$ and $\mathcal{H}^\dagger_{m,n}(\sigma)$ as families of statistics indexed by σ. Let us also note that for a fixed σ, one can compute $\mathcal{H}^\dagger_{m,n}(\sigma)$ and its p-value using Lemma 4.15 and Proposition 4.27.

We will first look at how the parameter σ affects the power of $\mathcal{H}_{m,n}(\sigma)$ in the univariate case, i.e. when $k = 1$. For the purpose, we will be interested in how σ affects the p-value $\mathbb{P}(\mathcal{H}_{m,n}(\sigma) > q)$ for a fixed $q > 0$, given $\boldsymbol{X}_m$ or $\boldsymbol{Y}_n$, the latter conditions being symmetric. Therefore in the sequel, we will be conditioning on $\boldsymbol{X}_m$. Define $q^*_{\boldsymbol{X}_m}(\sigma; p_0)$ as the conditional critical level

$$q^*_{\boldsymbol{X}_m}(\sigma; p_0) = \inf\{q : \mathbb{P}(\mathcal{H}_{m,n}(\sigma) \le q | \boldsymbol{X}_m) \ge p_0\}, \tag{4.4.1}$$

where p_0 is fixed. Let

$$\begin{aligned}
\bar{\mathcal{U}}_{F^c_m}(q,\sigma) &:= \left\{ (\frac{x-q}{\sigma}, z+q) : (x,z) \in \mathcal{U}_{F^c_{m,\sigma}} \right\} \equiv \left\{ (x - \frac{q}{\sigma}, z+q) : (x,z) \in \mathcal{U}_{F^c_m} \right\}, \\
\bar{\mathcal{L}}_{F^c_m}(q,\sigma) &:= \left\{ (\frac{x+q}{\sigma}, z-q) : (x,z) \in \mathcal{L}_{F^c_{m,\sigma}} \right\} \equiv \left\{ (x + \frac{q}{\sigma}, z-q) : (x,z) \in \mathcal{L}_{F^c_m} \right\},
\end{aligned} \tag{4.4.2}$$

which are defined by the domains $\mathcal{L}_{F^c_m}$ and $\mathcal{U}_{F^c_m}$ introduced in Section 4.2.2, appropriately modifying the coordinates x and z by q and σ. It is not difficult to see that the p-value of $\mathcal{H}_{m,n}(\sigma)$ conditional on $\boldsymbol{X}_m$ can be expressed as

$$\begin{aligned}
&\mathbb{P}\{\mathcal{H}_{m,n}(\sigma) \ge q | \boldsymbol{X}_m\} = \\
&\quad \mathbb{P}\left\{ B_j \in \bar{\mathcal{L}}_{F^c_m}(q,\sigma) \cup \bar{\mathcal{U}}_{F^c_m}(q,\sigma), \text{ for at least one } j \in \{1, \ldots, 2\nu\} \,\middle|\, \boldsymbol{X}_m \right\},
\end{aligned} \tag{4.4.3}$$

where the probability on the right-hand side can be interpreted as the probability that the ecdf G_n of $\boldsymbol{Y}_n$, with vertices B_j $j = 1, \ldots, 2\nu$ exits the corridor between the domains $\bar{\mathcal{U}}_{F_m^c}(q,\sigma)$ and $\bar{\mathcal{L}}_{F_m^c}(q,\sigma)$, which we denote by $\mathcal{M}_{F_m^c}(q,\sigma)$. The latter corridor is in fact the set that is complimentary to the set $\bar{\mathcal{L}}_{F_m^c}(q,\sigma) \cup \bar{\mathcal{U}}_{F_m^c}(q,\sigma)$. Therefore, based on (4.4.3), one can see that the p-value of H conditional on $\boldsymbol{X}_m$ depends on the shapes of $\bar{\mathcal{L}}_{F_m^c}(q,\sigma)$ and $\bar{\mathcal{U}}_{F_m^c}(q,\sigma)$ which depend on σ. Since we are interested in how change of scale affects the p-value, it would be useful to compare the shapes of $\bar{\mathcal{L}}_{F_m^c}(q^*_{\boldsymbol{X}_m}(\sigma;p_0),\sigma)$ and $\bar{\mathcal{U}}_{F_m^c}(q^*_{\boldsymbol{X}_m}(\sigma;p_0),\sigma)$ given $\boldsymbol{X}_m$ for a fixed p_0 and varying σ. For the purpose, we need to know $q^*_{\boldsymbol{X}_m}(\sigma)$, which is not directly available since the true distribution of $\boldsymbol{Y}_n$ and therefore of $\mathcal{H}_{m,n}(\sigma)$, is unknown. Nevertheless we can rely on the following proposition to investigate the behavior of $\bar{\mathcal{L}}_{F_m^c}(q^*_{\boldsymbol{X}_m}(\sigma;p_0),\sigma)$ and $\bar{\mathcal{U}}_{F_m^c}(q^*_{\boldsymbol{X}_m}(\sigma;p_0),\sigma)$ as σ varies.

Proposition 4.30. *Given $\boldsymbol{X_m}$, $\sigma_1 > \sigma_2 > 0$ and $p_0 > 0$, if $\boldsymbol{Y}_n$ comes from a continuous distribution, we have*

$$q^*_{\boldsymbol{X}_m}(\sigma_2;p_0) \leq q^*_{\boldsymbol{X}_m}(\sigma_1;p_0) \leq \frac{\sigma_1}{\sigma_2} q^*_{\boldsymbol{X}_m}(\sigma_2;p_0). \tag{4.4.4}$$

Therefore, in order to investigate the behavior of $\bar{\mathcal{L}}_{F_m^c}(q^*_{\boldsymbol{X}_m}(\sigma;p_0),\sigma)$ and $\bar{\mathcal{U}}_{F_m^c}(q^*_{\boldsymbol{X}_m}(\sigma;p_0),\sigma)$ when σ varies, we can select σ and a value $q^*(\sigma)$ for the unobservable $q^*_{\boldsymbol{X}_m}(\sigma)$ so that inequality (4.4.4) is satisfied. Applying this approach in Example 4.31, we illustrate graphically the dependence of $\bar{\mathcal{L}}_{F_m^c}(q^*(\sigma),\sigma)$ and $\bar{\mathcal{U}}_{F_m^c}(q^*(\sigma),\sigma)$ on σ, for a particular realization of $\boldsymbol{X}_m$.

Example 4.31. Let $\{0.1, 0.2, 0.4, 0.8, 1.6\}$ be a particular realization of $\boldsymbol{X}_m$. We select $\sigma_1 = 0.5$, $\sigma_2 = 1$, $\sigma_3 = 5$, $q^*(\sigma_1) = 0.07$, $q^*(\sigma_2) = 0.1$ and $q^*(\sigma_3) = 0.15$ which satisfy (4.4.4).

For the choice of values as in Example 4.31, in Figure 4.7, we present the corridor $\mathcal{M}_{F_m^c}(q^*(\sigma_i),\sigma_i)$ between $\bar{\mathcal{L}}_{F_m^c}(q^*(\sigma_i),\sigma_i)$ and $\bar{\mathcal{U}}_{F_m^c}(q^*(\sigma_i),\sigma_i)$ for $i = 1, 2, 3$. As can be seen, the smaller σ, the narrower (by area) the latter corridor is in the right tail. Consequently, a smaller σ increases the probability of capturing deviations in the right-tail of the distribution underlying $\boldsymbol{Y}_n$. In contrast, a larger σ leads to better capturing deviations in the left tail.

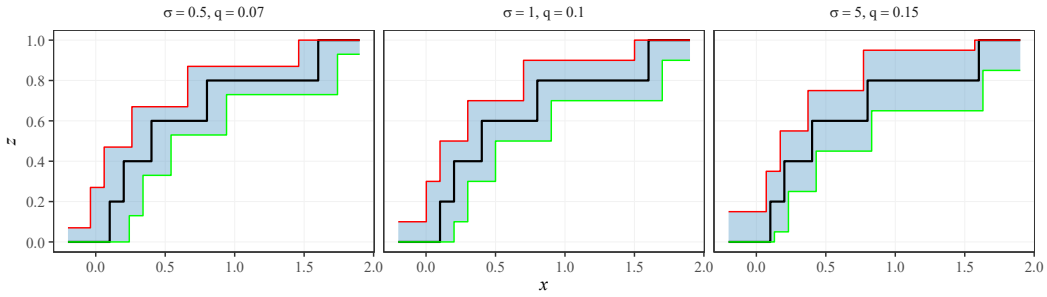


Figure 4.7: The corridor $\mathcal{M}_{F_m^c}(q^*(\sigma_i),\sigma_i)$ for different σ_i and $q(\sigma^*)$ in Example 4.31 given $\boldsymbol{X}_m = \{0.1, 0.2, 0.4, 0.8, 1.6\}$

.

Therefore, in general for a given $\boldsymbol{X}_m$, a rule for selecting σ analogue to (3.3.11) would be to standardize the distance between two quantiles, i.e.

$$\sigma = \frac{\psi_1 - \psi_2}{Q_{\boldsymbol{X}_m}(\psi_1) - Q_{\boldsymbol{X}_m}(\psi_2)}, \tag{4.4.5}$$

where $Q_{\boldsymbol{X}_m}(\psi) = X_{(\lceil m\psi \rceil)} + (m\psi - \lfloor m\psi \rfloor)X_{(\lfloor m\psi \rfloor)}$ is a smooth version of the ψ-th quantile of $\boldsymbol{X}_m$ $\psi \in (0,1)$, $X_{(i)}$ denotes the i-th order statistics of $\boldsymbol{X}_m$, $\lceil \cdot \rceil$ and $\lfloor \cdot \rfloor$ denotes rounding up and down, respectively.

Let us recall that as established by Theorems 4.24 and 4.25. the p-values of $\mathcal{H}_{m,n}(\sigma)$ and its permutation version $\mathcal{H}^{\dagger}_{m,n}(\sigma)$ are asymtotically equivalent and one can estimate $\mathbb{P}(\mathcal{H}_{m,n}(\sigma) > q)$ by computing $\mathbb{P}(\mathcal{H}^{\dagger}_{m,n}(\sigma) > q | \boldsymbol{Z}_{m+n})$ following Proposition 4.27 (cf. (4.3.14)).

However, using σ as of (4.4.5) to evaluate the p-value $\mathbb{P}(\mathcal{H}^{\dagger}_{m,n}(\sigma) > q | \boldsymbol{Z}_{m+n})$ would not be directly relevant as one would need to substitute $\boldsymbol{X}_m$ with $\boldsymbol{X}^{\dagger}_m$ in (4.4.5), which makes it dependent on the random vector $\boldsymbol{R}_{m+n}$. Another drawback of (4.4.5) is that it does not take into account the symmetry with respect to $\boldsymbol{X}_m$ and $\boldsymbol{Y}_n$ and only depends on $\boldsymbol{X}_m$. These problems are mitigated by the following expression for σ,

$$\sigma^* = \mathbb{E}_{\boldsymbol{R}_{m+n}}\left[\max\left(\frac{\psi_1 - \psi_2}{Q_{\boldsymbol{X}^{\dagger}_m}(\psi_1) - Q_{\boldsymbol{X}^{\dagger}_m}(\psi_2)}, \frac{\psi_1 - \psi_2}{Q_{\boldsymbol{Y}^{\dagger}_n}(\psi_1) - Q_{\boldsymbol{Y}^{\dagger}_n}(\psi_2)}\right) \middle| \boldsymbol{Z}_{m+n}\right], \tag{4.4.6}$$

where $\boldsymbol{X}^{\dagger}_m$ and $\boldsymbol{Y}^{\dagger}_n$ expressed through the random vector $\boldsymbol{R}_{m+n}$ are defined in Section 4.3.2 (cf. (4.3.9)) and $\psi_1, \psi_2 \in (0,1)$. As can be seen, the randomness with respect to $\boldsymbol{R}_{m+n}$ is eliminated and since (4.4.6) is conditional on $\boldsymbol{Z}_{m+n}$, it can be used to

evaluate $\mathbb{P}(\mathcal{H}^{\dagger}_{m,n}(\sigma) > q|\boldsymbol{Z}_{m+n})$ applying the exact expression (4.3.14). Similarly, (4.4.6) is also exact and computable either directly when m and n are small or by simulation when sample sizes are large.

Now we turn our attention to the choice of $\psi_1, \psi_2 \in (0,1)$. As shown in Sections 3.3 and 3.5, by appropriately selecting ψ_1 and ψ_2, one can control σ and therefore tune the one-sample H statistic to be sensitive in the left/right tail or in the body (see Table 3.1 therein). The domain of sensitivity of $\mathcal{H}_{m,n}$ can be similarly controlled by appropriately selecting ψ_1 and ψ_2 in (4.4.6) using the latter table. In order to ensure that $\mathcal{H}^{\dagger}_{m,n}$ is sensitive in either the left/right tail or in the body of $\mathcal{H}^{\dagger}_{m,n}$, it suffices for either $F^{\dagger}_m$ or $G^{\dagger}_n$ to exit $\mathcal{M}_{G^{\dagger c}_n}(q,\sigma)$ or $\mathcal{M}_{F^{\dagger c}_m}(q,\sigma)$ with higher probability respectively at this region. This can be guaranteed by making either $\mathcal{M}_{G^{\dagger c}_n}(q,\sigma)$ or $\mathcal{M}_{F^{\dagger c}_m}(q,\sigma)$ become narrower, which requires the maximum in (4.4.6).

We demonstrate in Example 4.33 that the two-sample $\mathcal{H}_{m,n}(\sigma)$ statistic becomes right-tail sensitive if ψ_1 and ψ_2 are set equal to 0.99 and 0.95 respectively, following Table 3.1.

Furthermore, in practice, the sample observations may be rescaled for different reasons e.g., using different scale coefficients, and presented in different units. It is not difficult to see that under the choice (4.4.6), $\mathcal{H}_{m,n}(\sigma^*)$ and its p-value will be invariant with respect to the units used to present the data, which are similarly stated in Proposition 3.35.

In the following theorem, we show that σ^* in (4.4.6) converges to the rule (3.3.11) in Section (3.3.2) with respect to the pooled distribution E.

Theorem 4.32. *Let $\boldsymbol{X}_m$ and $\boldsymbol{Y}_n$ have underlying unknown distribution F and G respectively. For $\psi_1 > \psi_2 \in (0,1)$, if $e(E^{-1}(\psi_1)), e(E^{-1}(\psi_2)) > 0$, where e and E^{-1} are the density and quantile functions of the pooled distribution E. Then we have*

$$\sigma^* \xrightarrow{\mathbb{P}} \sigma_0, \text{ as } m,n \to \infty, \frac{m}{m+n} \to \eta, \tag{4.4.7}$$

where $\sigma_0 = \frac{\psi_1 - \psi_2}{E^{-1}(\psi_1) - E^{-1}(\psi_2)}$ and σ^* is defined in (4.4.6). Furthermore, under the null or a fixed alternative as in Theorem 4.24, we have

$$q^{\dagger}_{m,n}(\sigma^*) - q^{*}_{m,n}(\sigma_0) \xrightarrow{\mathbb{P}} 0, \tag{4.4.8}$$

where $q^{\dagger}_{m,n}(\sigma^*)$ *and* $q^{*}_{m,n}(\sigma_0)$ *are correspondingly the critical values of* $\sqrt{\frac{mn}{m+n}}\mathcal{H}^{\dagger}_{m,n}(\sigma^*)$ *with* σ^* *defined in (4.4.6) and* $\sqrt{\frac{mn}{m+n}}\mathcal{H}_{m,n}(\sigma_0)$.

We will illustrate the application of the H test $\mathcal{H}^{\dagger}_{m,n}(\sigma^*)$ with σ^* selected according to (4.4.6) in the following example. A similar construction for the one-sample case is given in Example 3.44.

Example 4.33. We apply the splicing construction as in Example 3.43, where the alternative G has a density $g(x)$ defined by (3.4.9), satisfying conditions 1)-3). In other words, G coincides with F in the body and has the tail of G_0. In this construction, we assume that both F and G_0 come from the family of (unshifted) Fréchet distributions, which are characterized by scale and shape parameters α and θ. More precisely, the cdf of a Fréchet distribution is given as (see also (3.5.1))

$$\Phi(x;\alpha,\theta)=\begin{cases}\exp\Big(-(1+\frac{\theta x}{\alpha})^{-\frac{1}{\theta}}\Big) & x>-\frac{\alpha}{\theta}\\ 0 & x\le-\frac{\alpha}{\theta}\end{cases}.$$

In order to compare the powers of KS, CvM, AD, W, Kuiper and $\mathcal{H}^{\dagger}_{m,n}(\sigma)$, we take $F\sim F_{Fr}(x;0.3,0.3)$, $\underline{x}=-1$, $\phi_1=0.8$, $C=(-\log 0.8)^{-0.3}-1$, and $G_0\sim\Phi(x;\theta_0,\alpha_0)$. Since condition 3) ensures the uniqueness of α_0 for a fixed shape parameter θ_0, the choice of θ_0 fully determines the shape of G. In this example, we are interested in comparing numerically the powers of KS, CvM, AD, W, Kuiper and $\mathcal{H}^{\dagger}_{m,n}(\sigma)$, under different tail alternatives from the Fréchet family (3.5.1). For this reason, we choose $\theta_0=0.3+\Delta$, where $\Delta=0.2,0.4,0.6,\ldots,3.2$ reflects the tail difference between F and G. As shown in the left panel of Figure 4.8, with Δ increasing, the difference between G and F in the tail also increases.

Then we simulate 500 pairs of samples $\boldsymbol{X}_m$ and $\boldsymbol{Y}_n$ coming from $F(x)$ and $G(x)$ respectively, and compute the power as the frequency of rejecting the null for significance level $p=0.05$ applying the statistics KS, CvM, AD, W, Kuiper and $\mathcal{H}_{m,n}(\sigma)$. In order to fit the tail, for each pair of realizations of $\boldsymbol{X}_m$ and $\boldsymbol{Y}_n$, we estimate the optimal scale σ^* following (4.4.6) with $\psi_1=0.99$, $\psi_2=0.95$, based on 1000 simulations of the random vector $\boldsymbol{R}_{m+n}$ uniformly distributed over $\Pi(m+n)$.

The power of these statistics as a function of Δ is illustrated in the right panel of Figure 4.8, for sample sizes $m=n=50$, 100 and 300.

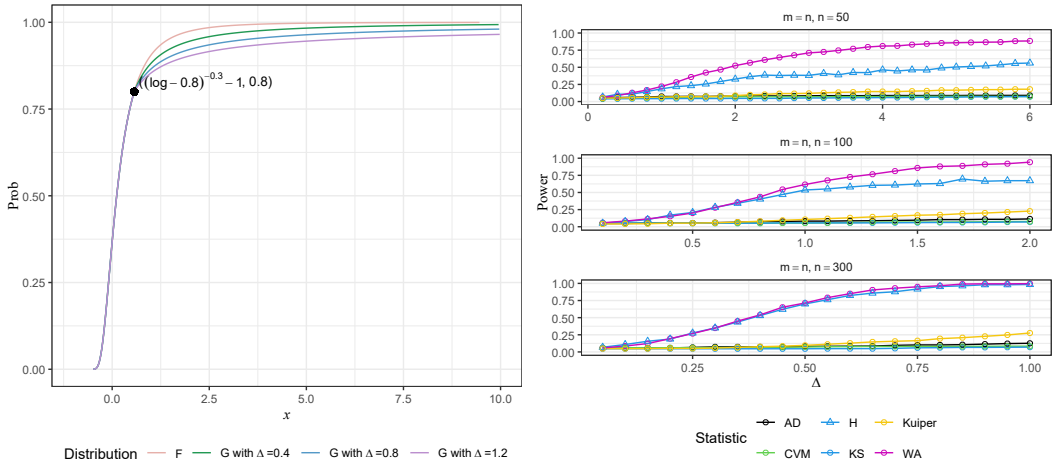


Figure 4.8: Left Panel: $F \sim \Phi(x; 0.3, 0.3)$ and G with tail $G_0 \sim \Phi(x; \theta_0, \alpha_0)$, where $\theta_0 = 0.3 + \Delta$; Right Panel: the powers of the KS, CvM, AD, W, Kuiper and $\mathcal{H}^{\dagger}_{m,n}(\sigma^*)$ tests, as functions of Δ.

Remark 4.34. If the proposed rule $\sigma = f(\boldsymbol{X}_m, \boldsymbol{Y}_n)$ happens to choose an extremely small scale for certain realizations, the Hausdorff test can become degenerate, i.e., its statistic shrinks and the resulting p-value type II error will be close to 1. When both samples are large, this abnormality is unlikely provided the selected scale converges to a positive limit σ_0. This is the case for the rule σ^*, which is guaranteed by the convergence result in Theorem 4.32. From this perspective, finer properties of the rule, such as unbiasedness, optimality in terms of variance, are less important, since they do not by themselves prevent degeneracy at small sample sizes.

As further illustrated in Example 3.44, the rule in (4.4.6) has the additional practical advantage that it keeps $\mathcal{H}^{\dagger}_{m,n}(\sigma^*)$ non-degenerate even when m and n are small.

As evidenced by Figure 4.8, the power of $\mathcal{H}^{\dagger}_{m,n}(\sigma^*)$ with σ^* chosen according to (4.4.6) is substantially higher than the power of all other tests except the Wasserstein (W), confirming the efficiency of $\mathcal{H}^{\dagger}_{m,n}(\sigma^*)$ in detecting tail differences. Furthermore, the power of $\mathcal{H}^{\dagger}_{m,n}(\sigma^*)$ tends to converge to the power of the W test when sample sizes increase, as can be seen from the right panel of Figure 4.8. This tendency is preserved for $m = n = 500$, the corresponding plot is very similar to that for $m = n = 300$ and is therefore omitted. We argue that this convergence stems from the shared geometric properties of H and W, both being area-based tests. More precisely, the univariate Wasserstein distance equals the integral between two empirical cdfs, while $\mathcal{H}^{\dagger}_{m,n}(\sigma^*)$

extracts only the side of the largest square that can be inserted between these cdfs. Although this square constitutes only a portion of the total integral, it contains the essential information for capturing tail differences, therefore replacing the need for full integration.

In the spirit of (4.4.6), for the multivariate case, we propose to choose $\sigma^* = (\sigma^{*(1)},\ldots,\sigma^{*(k)})^T$ following

$$\sigma^{*(i)} = \mathbb{E}\left[\max\left(\frac{\psi_1-\psi_2}{Q^{(i)}_{\boldsymbol{X}^\dagger_m}(\psi_1,\psi)-Q^{(i)}_{\boldsymbol{X}^\dagger_m}(\psi_2,\psi)}, \frac{\psi_1-\psi_2}{Q^{(i)}_{\boldsymbol{Y}^\dagger_n}(\psi_1,\psi)-Q^{(i)}_{\boldsymbol{Y}^\dagger_n}(\psi_2,\psi)}\right)\middle|\boldsymbol{Z}_{m+n}\right], \tag{4.4.9}$$

when $i=1,\ldots,k$, where $Q^{(i)}_{\boldsymbol{X}^\dagger_m}(\psi_1,\psi)$ is the ψ_1-quantile of the i-th component of $\boldsymbol{X}^\dagger_m$, conditional on the event that all its other components with index $j\neq i$ exceed their respective marginal ψ-quantiles. Clearly, when ψ is chosen to be close to 1, σ^* defined according to (4.4.9) captures the marginal shape of the tail in each direction.

We will illustrate the application of (4.4.9) in conjuction with the statistic $\mathcal{H}^\circ_{m,n}(\sigma^*)$ instead of $\mathcal{H}^\dagger_{m,n}(\sigma^*)$, since $\mathcal{H}^\circ_{m,n}(\sigma^*)$ is invariant with respect to the definitions $F_{m,i}$, $i=1,\ldots,2^k$, as can be seen from (4.2.17). The same definition invariance applies to a set of competing tests, including the Friedman and Rafsky (1979) (FR) run test, extented KS test by Peacock (1983), Ball Divergence (BD) test (Pan et al., 2018), Maximum Mean Discrepancy (MMD) test (Gretton et al., 2012) with Gaussian kernel, Cross Match (CM) test (Rosenbaum, 2005), Biswas et al. (2014) (BMG) run test, and Schilling-Henze Nearest Neighbor (NN) test (cf. Schilling, 1986; Henze, 1988). In Example 4.35, the power of the latter tests is compared with that of $\mathcal{H}^\circ_{m,n}(\sigma^*)$, where σ^* is selected according to (4.4.9). The required p-value of $\mathcal{H}^\circ_{m,n}(\sigma^*)$ is estimated by the p-value of its permutation version, as described in Section 4.3.2.

Example 4.35. We take $F_1 \sim EC(\mu_1,\Sigma_1,f)$ and $G_1 \sim EC(\mu_2,\Sigma_2,g)$ to be two elliptical distributions that are defined on $\mathbb{R}^2$ with density functions f_1 and g_1 respectively, i.e.

$$f_1 \propto f((x-\mu_1)^T\Sigma_1^{-1}(x-\mu_1)),\ g_1 \propto g((x-\mu_2)^T\Sigma_2^{-1}(x-\mu_2)),$$

where $\mu_1,\mu_2\in\mathbb{R}^2$, $\Sigma_1^{-1},\Sigma_2^{-1}$ are the inverse of the positive definite matrixes $\Sigma_1,\Sigma_2\in$

$\mathbb{R}^{2\times 2}$, f and g are positive integrable functions defined on $[0,\infty)$. Since we are interested in the performance of $\mathcal{H}^{\circ}_{m,n}(\sigma^*)$ when F and G are different in the tail, we require that $\mu_1=\mu_2$, $\Sigma_1=\Sigma_2$, $f(x)$ and $g(x)$ are univariate density functions meet the conditions 1)-3) in Example 3.43, i.e. $\int_0^\infty f_1(x)dx=\int_0^\infty g_1(x)dx=1$, $f_1(x)=g_1(x)$ up to a constant C, and differs from $g_1(x)$ when $x>C$.

We take $\mu_1=\mu_2=(0,0)^T$, $\Sigma_1=\Sigma_2=\begin{bmatrix}2 & 0.7\\ 0.7 & 2\end{bmatrix}$, $f=e^{-x}$, $\phi_1=0.5$, $C=-\log\phi_1=\log 2$, $g(x)=\begin{cases}e^{-x} & 0\le x\le C\\ \frac{\theta_0\alpha_0^{\theta_0}}{(x+\alpha_0)^{\theta_0+1}} & x>C\end{cases}$ with $\alpha_0=-\frac{\log\phi_1}{(1-\phi_1)^{-1/\theta_0}-1}$.

In this way, F is a multivariate Normal distribution and G is a spliced elliptical distribution with its center region coinciding with F. We have used the rejection sampling to draw $\boldsymbol{X}_m$ and $\boldsymbol{Y}_n$ correspondingly from F and G.

We evaluate σ^* as defined in (4.4.9) by setting $(\psi_1,\psi_2,\psi)=(0.99,0.95,0.7)$ and computing the corresponding quantiles using simulation. We then use σ^* to estimate the p-value of $\mathcal{H}^{\circ}_{m,n}(\sigma^*)$.

The above-mentioned power comparisons for particular choices of θ_0, α_0, m and n are summarized in Table 4.1. As can be seen, the power of $\mathcal{H}^{\circ}_{m,n}(\sigma)$ is mostly higher than the power of all the other tests except W. However, let us note that a

$m=n$	(θ_0,α_0)	$\mathcal{H}^{\circ}_{m,n}(\sigma^*)$	KS	BD	MMD	CM	BMG	FR	NN	W
50	(1.5,1.1800)	0.14	0.03	0.08	0.06	0.07	0.04	0.05	0.07	0.23
50	(2.0,1.6734)	0.07	0.02	0.04	0.05	0.04	0.06	0.06	0.07	0.22
100	(1.5,1.1800)	0.22	0.05	0.16	0.10	0.07	0.05	0.11	0.14	0.7
100	(2.0,1.6734)	0.11	0.04	0.14	0.07	0.08	0.03	0.07	0.11	0.43

Table 4.1: The powers of the extended KS, W, BD, MMD with Gaussian kernel, CM, BMG, FR, NN with $N=3$ and $\mathcal{H}^{\circ}_{m,n}(\sigma)$ with $\sigma=\sigma^*$ chosen following (4.4.9) when $m=n=50$ and 100, and some particular choice of (θ_0,α_0).

direct computation of W (without entropic regularization) becomes computationally demanding (see Cuturi, 2013), especially in higher dimensions.

4.5 Bahadur Exact Slope

As we have introduced in Section 3.4.2, the Bahadur exact slope of a statistic is a characterization for the exponential convergence rate of its p-value under a fixed alternative. The ratio of slopes of two different statistic, which is referred to as their Bahadur relative efficiency, also reflects their relative size of asymptotic power

(cf. (3.4.6)). In Theorem 3.41 of Section 3.4.2, we have provided an computable expression of the exact slope of the one-sample $\mathcal{H}_n(\sigma)$ test. In this section, we will provide some results to the exact slope of $\mathcal{H}_{m,n}(\sigma)$.

Similar to Section 3.4.2, we introduce the corresponding Bahadur exact slope of $\mathcal{H}_{m,n}(\sigma)$ as

$$c(\sigma;\eta) = -\lim_{m,n\to\infty,\frac{m}{m+n}\to\eta} 2(m+n)^{-1}\log P_{m,n}(\sigma) \quad a.s., \tag{4.5.1}$$

where $P_{m,n}(\sigma)$ is the random p-value of $\mathcal{H}_{m,n}(\sigma)$ under a fixed alternative $F \neq G$. In the following, we will generalize Theorem 3.41 and give to the exact slope of $\mathcal{H}_{m,n}$.

Theorem 4.36. *Let $\boldsymbol{X}_m$ and $\boldsymbol{Y}_n$ come from univariate distributions F and G, the exact slope defined in (4.5.1) exists, and can be expressed as $c(\sigma;\eta) = \frac{1}{2}f_{\sigma,\eta}(H(F_\sigma, G_\sigma))$, where F_σ and G_σ denote F and G scaled by σ, $f_{\sigma,\eta}(q) = -\max\{r_1(q), r_2(q)\}$, and where*

$$\begin{aligned}
r_1(q) &= \sup_{0<x<1}\left\{\eta\log(\varphi_x(\frac{\tau_1(x,q)}{\eta}) - q) + (1-\eta)\log(\varphi_x(\frac{-\tau_1(x,q)}{1-\eta})) - \tau_1(x,q)q\right\}\\
r_2(q) &= \sup_{0<x<1}\left\{\eta\log(\varphi_x(\frac{-\tau_1(x,q)}{\eta}) + q) + (1-\eta)\log(\varphi_x(\frac{\tau_1(x,q)}{1-\eta})) - \tau_2(x,q)q\right\}\\
\varphi_x(t) &= E(x)e^t + 1 - E(x)
\end{aligned} \tag{4.5.2}$$

and τ_1, τ_2 are functions of x and q implicitly defined by the following equations

$$\begin{aligned}
\varphi_x'(\frac{\tau_1}{\eta} - q)/\varphi_x(\frac{\tau_1}{\eta} - q) - \varphi_x'(\frac{\tau_1}{1-\eta})/\varphi_x(\frac{\tau_1}{1-\eta}) &= q\\
\varphi_x'(\frac{\tau_2}{\eta} + q)/\varphi_x(\frac{\tau_2}{\eta} + q) - \varphi_x'(\frac{\tau_2}{1-\eta})/\varphi_x(\frac{\tau_2}{1-\eta}) &= -q
\end{aligned} \tag{4.5.3}$$

4.6 Conclusions

We have proposed to use the two-sample Hausdorff (H) metric to measure the distance between two multivariate ecdfs, in the context of goodness-of-fit testing. The H test, which depends on both the ordinate and abscissa coordinates, is location invariant but scale dependent, in contrast to most of the classical tests, which are rank tests. The rank tests are computationally appealing since their p-values are independent of the underlying distribution. However, they have lower power when the samples have different tails. In addition, the rank tests are not continuous as

point functions of the observations, which affects the type I error. In contrast, the H test is Lipschitz continuous (see Theorem 4.6), which leads to its robustness to small perturbations in the sample observations. Furthermore, in Theorem 4.25 we show that H is qualitatively robust.

In Section 4.3.2, we consider the permutation version of the H test and show in Theorems 4.24, 4.25 and 4.26 that it is asymptotically equivalent to the original H test in terms of p-value, type I error, and power. This allows us to estimate the p-values of the H test which depend on the unknown null and therefore are not directly computable.

In Section 4.4, we have investigated the scale dependence of H, and have shown that, while such a property may seem to impede the use of H, it is, in fact, useful. It allows one to control locally the sensitivity of H and therefore optimize its power in the corresponding distributional domain (i.e. left/right tail or body). This is achieved by appropriately selecting the scale coefficient following Equation (4.4.6) or its multivariate generalization (4.4.9). We show by numerical examples (see Examples 4.33 and 4.35) that when the samples differ in the right tail, the scale-tuned univariate and bivariate H tests outperform many other tests in terms of power as summarized in Figure 4.8 and Table 4.1. In the univariate case, the power of H is lower than that of W for small sample sizes but tends to converge to the power of W for sample sizes equal to and exceeding $m = n = 300$, as illustrated in the right panel of Figure 4.8.

We have also obtained some additional results (cf. Proposition 4.9 and Theorem 4.19) that connect H to the Lévy-Prokhorov metric and the KS test. This offers a bridge to applying the H statistic in the context of global sensitivity analysis and variable importance measurement.

All this makes H a favorable alternative to other existing two-sample tests.

Appendix for Chapter 4

4.A Evaluating $\mathcal{H}_{m,n}$ when $k = 2$

This section is to give the method to compute $H(F_m^c, G_n^c)$ for the multivariate case when $k = 2$, as a further extension of the projection method in Section 4.3.1. Cases with $k > 2$ are not discussed here, as they involve only additional notational complications beyond the $k = 2$ case. For simplicity, we further assume that F and G are continuous, i.e. no tie is presented in the pooled sample with respect to each coordinate.

The task in this section is not trivial, since many existing results, which make the computation of $H(F_m^c, G_n^c)$ easier becomes complicated in the multivariate case. Firstly, the modification rule of the planar curves proposed in Section 4.2.4 is less meaningful for $k = 2$, since the calculation simplification benefit of introducing such a rule now becomes much less than the additional calculation cost raised due to complexity of the rule per se. Secondly, the projections of vertices, divide the real line into semi-open intervals in the case of $k = 1$, but no longer divide $\mathbb{R}^2$ as such when $k = 2$. Thirdly, set $\mathcal{G}$ in (4.2.21) is not bounded, thus, to implement Lemma 4.15, one needs to also incorporate Lemma 4.3. Given $x_1, \ldots, x_m \in \mathbb{R}^2$ and $y_1, \ldots, y_n \in \mathbb{R}^2$, the realizations of samples $\boldsymbol{X}_m$ and $\boldsymbol{Y}_n$ respectively, the computation of statisistics are as follows:

Step 1. Find the omnidirectional jumps of ecdfs and the vertices of F_m^c defined in (4.2.19). Let set $\mathcal{A}_{x1} = \{x_1^{(1)}, \ldots, x_m^{(1)}\}$ and $\mathcal{A}_{x2} = \{x_1^{(2)}, \ldots, x_m^{(2)}\}$. Then the omnidirectional jumps of F_m^c, $\alpha_1, \ldots, \alpha_\upsilon \in \mathbb{R}^2$, occur only on set $\mathcal{A}_{x1} \times \mathcal{A}_{x2}$. Hence, to find all the omnidirectional jumps, we could exhaustly check whether condition,

$$F_m(\alpha) \neq F_m(\alpha - \varepsilon_0 e_i), \text{for all } i = 1, 2 \tag{a}$$

holds, where $\alpha \in \mathcal{A}_{x1} \times \mathcal{A}_{x2}$, e_i is the basis vector in $\mathbb{R}^2$, ε_0 is a positive and sufficiently small number, and can take e.g.

$$\varepsilon_0 = \frac{1}{2} \min_{i=1,2; j \neq l} |x_j^{(i)} - x_l^{(i)}| \tag{4.A.1}$$

If condition (a) is satisfied, then α is an omnidirectional jump, corresponding to vertices $(\alpha, F_m(\alpha - \varepsilon))^T$ and $(\alpha, F_m(\alpha))^T$, where $\varepsilon \in \mathbb{R}^2$ are required to positive and sufficiently small with respect to each components, and can be taken as $\varepsilon = (\varepsilon_0, \varepsilon_0)^T$. Then, all the vertices in (4.2.19) of F_m^c that are used to compute $\mathcal{H}_{m,n}$ are $A_1 = (\alpha_1, F_m(\alpha_1 - \varepsilon))^T$, $A_2 = (\alpha_1, F_m(\alpha_1))^T$,..., $A_{2v-1} = (\alpha_v, F_m(\alpha_v - \varepsilon))^T$, $A_{2v} = (\alpha_v, F_m(\alpha_v))^T$. Similarly, we can find the omnidirectional jumps of G_n by checking condition (a) with respect to G_n on set $\mathcal{A}_{y1} \times \mathcal{A}_{y2}$, where set $\mathcal{A}_{y1} = \{y_1^{(1)}, \ldots, y_n^{(1)}\}$ and $\mathcal{A}_{y2} = \{y_1^{(2)}, \ldots, y_n^{(2)}\}$.

Step 2. Find the additional vertices of $\hat{F}_m^c(M)$, for an appropriately selected M. Note that to use Lemma 4.15, we need to incorporate Lemma 4.3 with appropriately selected M such that (4.2.10) holds, i.e. the metric between two truncated curves coincides with the value of the statistic. Thus, we select

$$M = 1 + \max\{x_j^{(i)}, y_l^{(i)}, 1 \leq i \leq k, 1 \leq j \leq m, 1 \leq l \leq n\}.$$

The additional vertices generated by truncation occur only on the set $(\mathcal{A}_{x1} \times \{M\}) \cup (\{M\} \times \mathcal{A}_{x2})$. Then we could check whether either one of the conditions

$$F_m(a, M) \neq F_m(a - \varepsilon_0, M), \tag{a*}$$

$$F_m(M, b) \neq F_m(M, b - \varepsilon_0) \tag{a**}$$

hold for $a \in \mathcal{A}_{x1}$ and $b \in \mathcal{A}_{x2}$ respectively. If condition (a*) is satisfied for $a \in \mathcal{A}_{x1}$, then we find its corresponding additional vertices $(a, M, F_m(a - \varepsilon_0, M))$ and $(a, M, F(a, M))$. Similarly, we could also find vertices $(M, b, F_m(M, b - \varepsilon_0))$ and $(M, b, F_m(M, b))$ through condition (a**). Finally, we fine all the boundary vertices

of $\hat{F}_m^c(M)$,

$$A_{2\upsilon+1}=(a_1,M,F_m(a_1-\varepsilon_0,M)),A_{2\upsilon+2}=(a_1,M,F_m(a_1,M)),\ldots,$$
$$A_{2(\upsilon+\upsilon')-1}=(a_{\upsilon'},M,F_m(a_{\upsilon'}-\varepsilon_0,M)),A_{2(\upsilon+\upsilon')}=(a_{\upsilon'},M,F_m(a_{\upsilon'},M)),$$
$$A_{2(\upsilon+\upsilon')+1}=(M,b_1,F_m(M,b_1-\varepsilon_0)),A_{2(\upsilon+\upsilon')+2}=(M,b_1,F_m(M,b_1)),\ldots,$$
$$A_{2(\upsilon+\upsilon'+\nu')-1}=(M,b_{\nu'},F_m(M,b_{\nu'}-\varepsilon_0)),A_{2(\upsilon+\upsilon'+\nu')}=(M,b_{\nu'},F_m(M,b_{\nu'}))$$

Step 3. Project the vertices A_l, $l=1,\ldots,2(\upsilon+\upsilon'+\nu')$ onto the $x^{(1)}x^{(2)}$ plane $\mathbb{R}^2$, along the direction of the vector $E_0=(1,1,-1)^T$ and denote these projections with λ_{A_l}. Clearly

$$\lambda_{A_l}=(w_{A_l}^{(1)}+w_{A_l}^{(3)},w_{A_l}^{(2)}+w_{A_l}^{(3)}),$$

where $w_{A_l}^{(i)}$, $i=1,2,3$ is the coordinates of A_l.

Step 4. Project all the vertices of G_n to $x^{(1)}x^{(2)}$ plane and constructing a partition of the $\mathbb{R}^2$ plane $x^{(1)}x^{(2)}$, similarly as Steps 3-4 in Section 4.2.4. Each jump

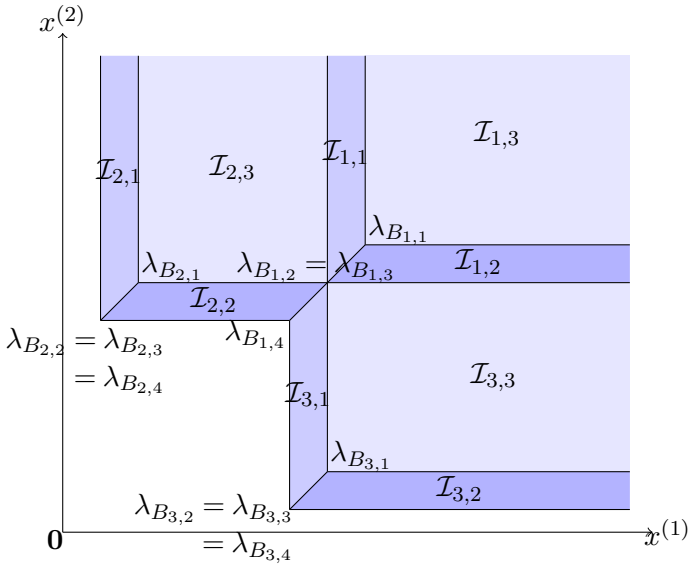


Figure 4.9: The vertex projection of G_n on plane $x^{(1)}x^{(2)}$, together with the faces projection.

β_i, $i=1,2,\ldots,\nu$, corresponds to 4 possible vertices, a locally concave vertex $B_{i,1}$, two saddle vertices $B_{i,2}$ and $B_{i,3}$ and a locally convex vertex $B_{i,4}$:

$$B_{i,1}=(\beta_i,G_n(\beta_i))^T,\ B_{i,2}=\left(w_{\beta_i}^{(1)},w_{\beta_i}^{(2)},G_n(w_{\beta_i}^{(1)}-\varepsilon_0',w_{\beta_i}^{(2)})\right)^T,$$
$$B_{i,3}=\left(w_{\beta_i}^{(1)},w_{\beta_i}^{(2)},G_n(w_{\beta_i}^{(1)},w_{\beta_i}^{(2)}-\varepsilon_0')\right)^T,\ B_{i,4}=\left(w_{\beta_i}^{(1)},w_{\beta_i}^{(2)},G_n(w_{\beta_i}^{(1)}-\varepsilon_0',w_{\beta_i}^{(2)}-\varepsilon_0')\right)^T,$$

where ε_0' is obtained via (4.A.1) by substituting y for x. The two saddle vertices and the locally convex may coincide. If there is no tie on both coordinates, the two saddle vertices always coincide, in which case we only refer to the vertex $B_{i,2}$. We project all the vertices as in Step 3, together with their surfaces, as illustrated in Figure 4.9. For clarity in presenting the algorithm, we further assign conceptual labels to the projection of the surfaces, though these notations are purely illustrative. Specifically, the projected regions $\mathcal{I}_{i,1}$, $\mathcal{I}_{i,2}$ and $\mathcal{I}_{i,3}$ use a two-index label, the first aligns with that of the locally concave vertices $B_{i,1}$, and the second reflects the normal vector direction of the corresponding surfaces, as in Figure 4.9. Alternatively, one can also view that the locally concave vertices $B_{i,1}$ would generate three projected surfaces. Therefore,

$$\mathcal{I}_{1,1},\mathcal{I}_{1,2},\mathcal{I}_{1,3},\ldots,\mathcal{I}_{\nu,1},\mathcal{I}_{\nu,2},\mathcal{I}_{\nu,3},\mathbb{R}^2/(\bigcup_{1\le i\le\nu,1\le s\le 3}\mathcal{I}_{i,s}) \tag{4.A.2}$$

forms a partition of $\mathbb{R}^2$. In addition, we also give the following lemma, describing the relative positions of omnidirectional jumps and projections,

Lemma 4.37. *For any two omnidirectional jumps β_{j_0} and β_{j_1} of G_n, if they are discordant, i.e. $\operatorname{sgn}(\beta_{j_0}^{(1)}-\beta_{j_1}^{(1)})=-\operatorname{sgn}(\beta_{j_0}^{(2)}-\beta_{j_1}^{(2)})$, there must exist an omnidirectional jump of G_n*

$$\beta^{max}=(\max(\beta_{j_0}^{(1)},\beta_{j_1}^{(2)}),\max(\beta_{j_0}^{(2)},\beta_{j_1}^{(2)}))^T.$$

Furthermore, for any two discordant projections, $\lambda_{B_{j_0,t_0}}$ and $\lambda_{B_{j_1,t_1}}$, $t_0,t_1\in\{1,2\}$ then there must exists a projection $\lambda_{B_{j_2,t_2}}\preceq(\max(\lambda_{B_{j_0,t_0}}^{(1)},\lambda_{B_{j_1,t_1}}^{(1)}),\max(\lambda_{B_{j_0,t_0}}^{(2)},\lambda_{B_{j_1,t_1}}^{(2)}))^T$ with $t_2\in\{2,4\}$ such that $\lambda_{B_{j_2,t_2}}\succeq\lambda_{B_{j_0,t_0}}$ or $\lambda_{B_{j_2,t_2}}\succeq\lambda_{B_{j_1,t_1}}$ holds.

For convenience, we define $\lambda_{B_{j_2,2}}$ as the successor of both $\lambda_{B_{j_0,1}}$ and $\lambda_{B_{j_1,1}}$, and $\lambda_{B_{j_2,4}}$ as the successor of both $\lambda_{B_{j_0,2}}$ and $\lambda_{B_{j_1,2}}$. Conversely, $\lambda_{B_{j_0,1}}$ and $\lambda_{B_{j_1,1}}$ are the predecessors of $\lambda_{B_{j_2,2}}$, while $\lambda_{B_{j_0,2}}$ and $\lambda_{B_{j_1,2}}$ are the predecessors of $\lambda_{B_{j_2,4}}$. For instance, in Figure 4.9, the successor of $\lambda_{B_{2,1}}$ and $\lambda_{B_{3,1}}$ is $\lambda_{B_{1,1}}$, and $\lambda_{B_{1,4}}$ is the predecessor of $\lambda_{B_{2,2}}$ and $\lambda_{B_{3,2}}$.

Step 5. Compute the distance $\inf_{B\in G_n^c}\rho_\infty(A_l,B)$ for some $l=1,\ldots,2(\upsilon+\upsilon'+\nu')$ and then compute $H(F_m^c,G_n^c)$. Note that in this step, to compute $H(F_m^c,G_n^c)$, one do not need to compute $\inf_{B\in G_n^c}\rho_\infty(A_l,B)$ for each $l=1,\ldots,2(\upsilon+\upsilon'+\nu')$. In fact, this result can be further simplified in the spirit of Lemma 4.18. In Lemma 4.18, all

the even vertices that are used to compute the distance are also the vertices that have the locally furthest distance. Therefore, we only need to compute $\inf_{B\in G_n^c}\rho_\infty(A_l,B)$ if A_l is locally furthest to G_n^c, i.e. $A_l\in\mathcal{V}_{loc}$, where

$$\mathcal{V}_{loc}=\left\{A_l\in\begin{cases}\mathcal{U}_{G_n^c} & \text{and } l \text{ is odd}\\ \mathcal{L}_{G_n^c} & \text{and } l \text{ is even}\end{cases} : l=1,\dots,2(\upsilon+\upsilon'+\nu')\right\}$$

.

Then, given an arbitrary projection $\lambda_{A_l}\in\mathcal{V}_{loc}$, our goal is to compute the distance by (4.2.21). Ideally, by Lemma 4.15, one should explicitly find the intersection E_l of line $\mathcal{L}_l$ defined according to (4.2.20) and the planar curve G_n^c and compute $\inf_{B\in G_n^c}\rho_\infty(A_l,B)=\rho_\infty(A_l,E_l)$. However, in the spirit of Lemma 4.23, it suffices to compute $\inf_{B\in G_n^c}\rho_\infty(A_l,B)$ if we locate which region $\mathcal{I}_{j,t}$ the projection λ_{A_l} falls into.

In order to find which set in (4.A.2) λ_{A_l} falls into, we first find the vertices projections of G_n^c satisfying

$$\mathcal{V}(\lambda_{A_l})=\{\lambda_{B_{j,t}}:\lambda_{A_l}-\lambda_{B_{j,t}}\succeq 0,\ j=1,\dots,\nu\ ,t=1,2,3,4\}.$$

This is a useful subset since, we expect that there exists $\lambda_{B_j,t_0}\in\mathcal{V}(\lambda_{A_l})$ such that $\lambda_{B_j,t_0},\lambda_{A_l}\in\mathcal{I}_{j,t}$ for some j and t. However, set $\mathcal{V}$ is quite large. To improve the searching efficiency, we further consider its subset consists of all the frontier projections of $\mathcal{V}(\lambda_{A_l})$, i.e.

$$\tilde{\mathcal{V}}(\lambda_{A_l})=\{\lambda_{B_{j_0,t_0}}\in\mathcal{V}(\lambda_{A_l}):\forall\ \lambda_{B_{j,t}}\neq\lambda_{B_{j_0,t_0}}\in\mathcal{V}(\lambda_{A_l}),\lambda_{B_{j_0,t_0}}\not\succeq\lambda_{B_{j,t}}\}.$$

On the one hand, since all the region we have considered in (4.A.2) does not overlap each other, the set $\tilde{\mathcal{V}}(\lambda_{A_l})$ still keeps the projection λ_{B_j,t_0} such that $\lambda_{B_j,t_0},\lambda_{A_l}\in\mathcal{I}_{j,t}$. On the other hand, as a direct consequence Lemma 4.37, only one projection of Type 1 or 2 could fall in $\tilde{\mathcal{V}}(\cdot)$, as formally stated in Proposition 4.38.

Proposition 4.38. *For any $(x,y)\in\mathbb{R}^2$, there does not exists $1\le j_0\neq j_1\le\nu$, such that $\lambda_{B_{j_0,t_0}},\lambda_{B_{j_1,t_1}}\in\tilde{\mathcal{V}}(x,y)$, where $t_0,t_1\in\{1,2\}$.*

Hence, from the above proposition, the set $\tilde{\mathcal{V}}(\lambda_{A_l})$ would only be one of the

following 3 cases.

- Case 1. $\tilde{\mathcal{V}}(\lambda_{A_l})$ is empty. Then $\lambda_{A_l} \in \mathbb{R}^2/(\bigcup \mathcal{I}_{i,s})$, thus $\inf_{B \in G_n^c} \rho_\infty(A_l, B) = |w_{A_l}^{(3)}|$
- Case 2. There exists j_0, such that $\lambda_{B_{j_0,t_0}} \in \tilde{\mathcal{V}}(\lambda_{A_l})$, where $t_0 = 1$ or 2.
- Case 3. $\tilde{\mathcal{V}}(\lambda_{A_l})$ is not empty and $\forall\, j = 1, \ldots, \nu$ and $t = 1, 2$, $\lambda_{B_{j,t}} \notin \tilde{\mathcal{V}}(\lambda_{A_l})$.

In Case 2 and 3, to compute $\inf_{B \in G_n^c} \rho_\infty(A_l, B)$, we need to further identify the projection $\lambda_B^*(A_l)$ sharing the same region in (4.A.2) with λ_{A_l}. In Case 2, we should have

$$\lambda_B^*(A_l) = \lambda_{B_{j_0,t_0}},\ t_0 = 1, 2 \text{ where } \lambda_{B_{j_0,t_0}} \in \tilde{\mathcal{V}}(\lambda_{A_l}) \tag{4.A.3}$$

In order to compute $\inf_{B \in G_n^c} \rho_\infty(A_l, B)$, we additionally require knowledge about the normal direction of the surface corresponding to the projected region $\mathcal{I}_{j,t}$ that contains A_l and $\lambda_B^*(A_l)$, i.e. the second subscript t of $\mathcal{I}_{j,t}$. This can be identified by the information of $\lambda_B^*(A_l)$. When the representative element $\lambda_B^*(A_l) = \lambda_{B_{j_0,1}}$ for some j_0, it is clear that both $A_l, B_{j_0,1} \in \mathcal{I}_{j_0,3}$. When $\lambda_B^*(A_l) = \lambda_{B_{j_0,2}}$, both $\lambda_{A_l}, \lambda_{B_{j_0,2}} \in \begin{cases} \mathcal{I}_{j_0,1} & \text{if } w_{B_{j_0,2}}^{(1)} - w_{A_l}^{(1)} \le w_{B_{j_0,2}}^{(2)} - w_{A_l}^{(2)} \\ \mathcal{I}_{j_0,2} & \text{if } w_{B_{j_0,2}}^{(1)} - w_{A_l}^{(1)} > w_{B_{j_0,2}}^{(2)} - w_{A_l}^{(2)}. \end{cases}$

In Case 3, we should have

$$\lambda_B^*(A_l) = \underset{\lambda_{B_{j,t}} \in \tilde{\mathcal{V}}(\lambda_{A_l})}{\arg\min}\ \rho_\infty^*(\lambda_{B_{j,t}}, \lambda_{A_l}). \tag{4.A.4}$$

However, it is not sufficient to compute $\inf_{B \in G_n^c} \rho_\infty(A_l, B)$, since we are not clear about the normal direction of $\mathcal{I}_{j,t}$ containing $\lambda_B^*(A_l)$. To obtain this, we need to further find the predecessors $\lambda_{B_{j_1,2}}$ and $\lambda_{B_{j_2,2}}$ in $\mathcal{V}(\lambda_{A_l})$ of $\lambda_B^*(A_l)$ in (4.A.4). Since $\mathcal{I}_{j,t}$ contains $\lambda_B^*(A_l)$, one of the predecessors $\lambda_{B_{j_1,2}}$ and $\lambda_{B_{j_2,2}}$ must fall in $\lambda_B^*(A_l)$. Such a predecessor is found by

$$\lambda_B^*(A_l) = \underset{\lambda \in \left\{\lambda_{B_{j_1,2}}, \lambda_{B_{j_2,2}}\right\}}{\arg\min}\ \rho_\infty(\lambda_{A_l}, \lambda). \tag{4.A.5}$$

Thus, identifying the normal direction of the region $\mathcal{I}_{j,t}$ will be the same as in (4.A.3).

In summary, we are able to compute $\inf_{B \in G_n^c} \rho_\infty(A_l, B)$ by the following lemma.

Lemma 4.39. *When $\lambda_B^*(A_l)$ is determined by (4.A.3) or (4.A.5), and $\lambda_B^*(A_l) = \lambda_{B_{j_0,t_0}}$ for some $1 \le j_0 \le \nu$ and $t_0 = 1$ or 2, then*

$$\inf_{B\in G_n^c} \rho_\infty(A_l, B) = \begin{cases} |w_{A_l}^{(1)} - w_{B_{j_0,t_0}}^{(1)}|, & \text{if } t_0 \neq 1 \text{ and } w_{B_{j_0,2}}^{(1)} - w_{A_l}^{(1)} \le w_{B_{j_0,2}}^{(2)} - w_{A_l}^{(2)} \\ |w_{A_l}^{(2)} - w_{B_{j_0,t_0}}^{(2)}|, & \text{if } t_0 \neq 1 \text{ and } w_{B_{j_0,2}}^{(1)} - w_{A_l}^{(1)} > w_{B_{j_0,2}}^{(2)} - w_{A_l}^{(2)} \\ |w_{A_l}^{(3)} - w_{B_{j_0,t_0}}^{(3)}|, & \text{if } t_0 = 1 \end{cases} \tag{4.A.6}$$

Eventually, $H(F_m^c, G_n^c) = \max\limits_{A_l\in\mathcal{V}_{loc}} \inf\limits_{B\in G_n^c} \rho_\infty(A_l, B)$. This finishes the description of the required algorithm. To compute other alternatives, $H(F_{m,i}^c, G_{n,i}^c)$ $i = 2, 3, 4$, one only needs to apply the same procedure above on both samples, using their corresponding coordinates with opposite signs.

4.B The Expression of the p-values as a Boundary Crossing Problem

Let us note that $\mathbb{P}(\mathcal{H}_{m,n}^\dagger \le q)$ can be expressed in terms of the probability of a trajectory $G_n^{\dagger c}$ staying within a corridor between two boundaries. The detailed computations are given in Appendix 4.B.

To see this, we give the result for $k = 1$ case and consider the set of points $(x, z) \in \mathbb{R}^2$ such that $\inf_{A\in F_m^{\dagger c}} \rho_\infty(A, (x, z)) \le q$ which defines a corridor around the curve $F_m^{\dagger c}$.

Define the set $\bar{\mathcal{U}}_{F_m^{\dagger c}} = \{(x-q, z+q) : (x, z) \in \mathcal{U}_{F_m^{\dagger c}}\}$ (The open set above in Figure 4.10). Similarly, define the set $\bar{\mathcal{L}}_{F_m^{\dagger c}} = \{(x-q, z+q) : (x, z) \in \mathcal{L}_{F_m^{\dagger c}}\}$. Thus, the set $\bar{\mathcal{U}}_{F_m^{\dagger c}} \cup \bar{\mathcal{L}}_{F_m^{\dagger c}} = \mathbb{R}^2 / \{(x, z) : \inf_{A\in F_m^{\dagger c}} \rho_\infty(A, (x, z)) \le q\}$. Therefore, one can see that the event $\{H(F_m^c, G_n^c) > q\}$ is equivalent to the event that at least one of the vertices $B_j^\dagger$ of $G_n^{\dagger c}$ falls in the region $\bar{\mathcal{U}}_{F_m^{\dagger c}} \cup \bar{\mathcal{L}}_{F_m^{\dagger c}}$ i.e.,

$$\begin{aligned} &\{\mathcal{H}_{m,n}^\dagger > q | \boldsymbol{R}_{m+n} = \pi_i(m+n), \boldsymbol{Z}_{m+n}\} \\ &\equiv \{B_j^\dagger \in \bar{\mathcal{U}}_{F_m^{\dagger c}} \cup \bar{\mathcal{L}}_{F_m^{\dagger c}}, \text{ for at least one } j \in \{1, \ldots, 2\nu\} | \boldsymbol{R}_{m+n} = \pi_i(m+n), \boldsymbol{Z}_{m+n}\} \end{aligned} \tag{4.B.1}$$

From (4.B.1) it follows that,

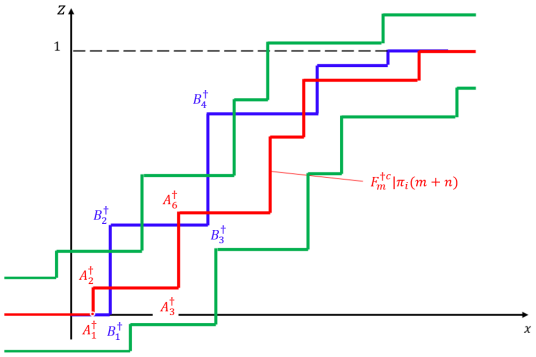


Figure 4.10: The probability $\mathbb{P}(\mathcal{H}_{m,n} \leq q)$ as a double boundary crossing problem

$$\begin{aligned}&\mathbb{1}\{\mathcal{H}^\dagger_{m,n} \leq q | \boldsymbol{R}_{m+n} = \pi_i(m+n), \boldsymbol{Z}_{m+n}\}\\ &= 1 - \mathbb{1}\{B_j^\dagger \in \bar{\mathcal{U}}_{F_m^{\dagger c}} \cup \bar{\mathcal{L}}_{F_m^{\dagger c}}, \text{ for at least one } j \in \{1,\ldots,2\nu\} | \boldsymbol{R}_{m+n} = \pi_i(m+n), \boldsymbol{Z}_{m+n}\}\end{aligned} \tag{4.B.2}$$

Substituting (4.B.2) in (4.3.14) we obtain that,

$$\begin{aligned}&\mathbb{P}(\mathcal{H}^\dagger_{m,n} \leq q | \boldsymbol{Z}_{m+n})\\ &= 1 - \frac{1}{C}\sum_{i=1}^{C} \mathbb{1}\Big\{B_j^\dagger \in \bar{\mathcal{U}}_{F_m^{\dagger c}} \cup \bar{\mathcal{L}}_{F_m^{\dagger c}},\\ &\qquad \text{for at least one } j \in \{1,\ldots,2\nu\} | \boldsymbol{R}_{m+n} = \pi_i(m+n), \boldsymbol{Z}_{m+n}\Big\}\end{aligned} \tag{4.B.3}$$

4.C Proofs for Chapter 4

This appendix provides proofs to all the statements in Chapter 3. Before we present some important results, it would be useful to give some auxiliary results and properties first.

We will start with Lemma 4.40, which is a multivariate generalization of Lemma 3.48.

Lemma 4.40. *For any $A \in \mathbb{R}^{k+1}$ and any planar curve F^c of F that is monotonic with respect to $\preceq$, the infimum $\inf_{B\in F^c} \rho_\infty(A,B)$ is always attained. Furthermore, if $\inf_{B\in F^c} \rho_\infty(A,B) > 0$, let B_0 be the crossing point of F^c and the line passing through A parallel to the vector OE_0 where $E_0 = (1,\ldots,1,-1)^T \in \mathbb{R}^{k+1}$, then $\rho_\infty(A,B_0) = \inf_{B\in F^c} \rho_\infty(A,B)$.*

Proof. Note that the planar curve defined by (4.2.6) is closed, therefore for an arbitrary ε the set $F^c \cap S(A, \inf_{B\in F^c} \rho_\infty(A,B) + \varepsilon)$ is compact. Hence, $\inf_{B\in F^c} \rho_\infty(A,B)$ is obtainable.

In addition, since the vector OE_0 is parallel to the vector AB_0, we have $\rho_\infty(A,B_0)=|w_A^{(i)}-w_{B_0}^{(i)}|$ for $i=1,\ldots,k+1$, where $w_A^{(i)},w_{B_0}^{(i)}$ are the i-th component of the point A and B_0 respectively. Without loss of genreality, we assume that $w_A^{(k+1)}>w_{B_0}^{(k+1)}$. This implies that $w_A^{(i)}<w_{B_0}^{(i)}$. For any $B\in\mathbb{R}^{k+1}$, if $\rho_\infty(A,B)<\rho_\infty(A,B_0)$, we must have $|w_A^{(i)}-w_B^{(i)}|<\rho_\infty(A,B_0)$ for every $i=1,\ldots,k+1$, where $w_{B_0}^{(i)}$ is the i-th component of the point B. For every $x\in\mathbb{R}^k$, when $x\preceq x_{B_0}=(w_{B_0}^{(1)},\ldots,w_{B_0}^{(k)})$, due to the monotonicity of F, we have $F(x)\leq w_{B_0}^{(k+1)}$. Therefore, if $x\preceq x_{B_0}$, $w_A^{(k+1)}-F(x+)\geq\rho_\infty(A,B_0)$. Then any point B such that $|w_A^{(i)}-w_B^{(i)}|<\rho_\infty(A,B_0)$ for every $i=1,\ldots,k+1$, i.e. $\rho_\infty(A,B)<\rho_\infty(A,B_0)$, we have $B\notin F^c$. □

4.C.1 Main Results in Chapter 4

Proof of Lemma 4.3. Let us take

$$C=\max_{\substack{1\leq i\leq k\\ 1\leq j\leq m,1\leq l\leq n}}\{X_j^{(i)},Y_l^{(i)}\}.$$

Then for any $x\in\mathbb{R}^k$, if $x^{(i)}>C$, then it is easy to show that $F_m(x)=F_m(x+e_i\varepsilon)$ and $G_n(x)=G_n(x+e_i\varepsilon)$ for any $\varepsilon>0$, where $e_i=(0,\ldots,0,1,0,\ldots,0)^T\in\mathbb{R}^k$ is the i-th standard basis vector of $\mathbb{R}^k$, with 1 appears exactly in the i-th position. Therefore, if $A\in F_m^c\cap\big[(\mathbb{R}^k-(-\infty,C]^k)\times[0,1]\big]$ and $B\in G_n^c\cap\big[(\mathbb{R}^k-(-\infty,C]^k)\times[0,1]\big]$,

$$A+e_i^*t\in F_m^c\cap\big[(\mathbb{R}^k-(-\infty,C]^k)\times[0,1]\big]\text{ and }B+e_i^*t\in G_n^c\cap\big[(\mathbb{R}^k-(-\infty,C]^k)\times[0,1]\big]\tag{4.C.1}$$

for any $t>0$ and $i=1,\ldots,k$, where $e_i^*=(e_i,0)^T\in\mathbb{R}^{k+1}$. Let us decompose both F_m^c and G_n^c into two parts,

$$\begin{aligned}&F_m^c\cap(-\infty,C]^k\times[0,1]\text{ and }F_m^c\cap\big[(\mathbb{R}^k-(-\infty,C]^k)\times[0,1]\big],\\&G_n^c\cap(-\infty,C]^k\times[0,1]\text{ and }G_n^c\cap\big[(\mathbb{R}^k-(-\infty,C]^k)\times[0,1]\big].\end{aligned}$$

For simplicity, we denote

$$\begin{aligned}&F_m^c(C)=F_m^c\cap(-\infty,C]^k\times[0,1],\ F_m^c(C^-)=\overline{F_m^c\cap\big[(\mathbb{R}^k-(-\infty,C]^k)\times[0,1]\big]},\\&G_n^c(C)=G_n^c\cap(-\infty,C]^k\times[0,1],\ G_n^c(C^-)=\overline{G_n^c\cap\big[(\mathbb{R}^k-(-\infty,C]^k)\times[0,1]\big]}.\end{aligned}\tag{4.C.2}$$

Clearly, for a fixed M

$$H(\hat{F}_m^c(M),\hat{G}_n^c(M)) = \max\{\sup_{A\in F_m^c(M)}\inf_{B\in G_n^c(M)}\rho_\infty(A,B), \sup_{A\in G_n^c(M)}\inf_{B\in F_m^c(M)}\rho_\infty(A,B)\}.$$

In order to show (4.2.10), it suffices to show that $\sup_{A\in F_m^c(M)}\inf_{B\in G_n^c(M)}\rho_\infty(A,B) = \sup_{A\in F_m^c}\inf_{B\in G_n^c}\rho_\infty(A,B)$ and $\sup_{A\in G_n^c(M)}\inf_{B\in F_m^c(M)}\rho_\infty(A,B) = \sup_{A\in G_n^c}\inf_{B\in F_m^c}\rho_\infty(A,B)$ for a sufficiently large M.

In order to show this, we first note that

$$\sup_{A\in F_m^c}\inf_{B\in G_n^c}\rho_\infty(A,B) = \max\Big\{\min\{\sup_{A\in F_m^c(C)}\inf_{B\in G_n^c(C)}\rho_\infty(A,B), \sup_{A\in F_m^c(C)}\inf_{B\in G_n^c(C^-)}\rho_\infty(A,B)\},$$
$$\min\{\sup_{A\in F_m^c(C^-)}\inf_{B\in G_n^c(C)}\rho_\infty(A,B), \sup_{A\in F_m^c(C^-)}\inf_{B\in G_n^c(C^-)}\rho_\infty(A,B)\}\Big\}.$$

We can easily verify that the term $\sup_{A\in F_m^c(C)}\inf_{B\in G_n^c(C^-)}\rho_\infty(A,B)$ and $\sup_{A\in F_m^c(C^-)}\inf_{B\in G_n^c(C)}\rho_\infty(A,B)$ diverges. Thus,

$$\sup_{A\in F_m^c}\inf_{B\in G_n^c}\rho_\infty(A,B) = \max\Big\{\sup_{A\in F_m^c(C)}\inf_{B\in G_n^c(C)}\rho_\infty(A,B), \sup_{A\in F_m^c(C^-)}\inf_{B\in G_n^c(C^-)}\rho_\infty(A,B)\Big\}.$$

Then we take $M_0 = C+1$, for any $M \geq M_0$,

$$\sup_{A\in F_m^c(M)}\inf_{B\in G_n^c(M)}\rho_\infty(A,B) = \max\Big\{\min\{\sup_{A\in F_m^c(C)}\inf_{B\in G_n^c(C)}\rho_\infty(A,B),$$
$$\sup_{A\in F_m^c(C)}\inf_{B\in \overline{G_n^c(M)-G_n^c(C^-)}}\rho_\infty(A,B)\},$$
$$\min\{\sup_{A\in \overline{F_m^c(M)-F_m^c(C^-)}}\inf_{B\in G_n^c(C)}\rho_\infty(A,B),$$
$$\sup_{A\in \overline{F_m^c(M)-F_m^c(C^-)}}\inf_{B\in \overline{G_n^c(M)-G_n^c(C^-)}}\rho_\infty(A,B)\}\Big\}.$$

Since

$$\sup_{A\in F_m^c(C)}\inf_{B\in G_n^c(C)}\rho_\infty(A,B) < 1, \quad \sup_{A\in \overline{F_m^c(M)-F_m^c(C^-)}}\inf_{B\in \overline{G_n^c(M)-G_n^c(C^-)}}\rho_\infty(A,B) < 1,$$
$$\sup_{A\in F_m^c(C)}\inf_{B\in \overline{G_n^c(M)-G_n^c(C^-)}}\rho_\infty(A,B) \geq 1, \quad \sup_{A\in \overline{F_m^c(M)-F_m^c(C^-)}}\inf_{B\in G_n^c(C)}\rho_\infty(A,B) \geq 1,$$

we have

$$\sup_{A\in F_m^c(M)}\inf_{B\in G_n^c(M)}\rho_\infty(A,B)=\max\Big\{\sup_{A\in F_m^c(C)}\inf_{B\in G_n^c(C)}\rho_\infty(A,B),\ \sup_{A\in\overline{F_m^c(M)-F_m^c(C^-)}}\inf_{B\in\overline{G_n^c(M)-G_n^c(C^-)}}\rho_\infty(A,B)\}\Big\}.$$

Both $\sup_{A\in F_m^c(C)}\inf_{B\in G_n^c(C)}\rho_\infty(A,B)$ and $\sup_{A\in\overline{F_m^c(M)-F_m^c(C^-)}}\inf_{B\in\overline{G_n^c(M)-G_n^c(C^-)}}\rho_\infty(A,B)$ are obtainable, since both $F_m^c(M)\triangle G_n^c(M)$ and $(F_m^c(M)-F_m^c(C^-))\triangle(G_n^c(M)-G_n^c(C^-))$ are bounded.

Then to show $\sup_{A\in F_m^c(M)}\inf_{B\in G_n^c(M)}\rho_\infty(A,B)=\sup_{A\in F_m^c}\inf_{B\in G_n^c}\rho_\infty(A,B)$, it remains to show that

$$\sup_{A\in F_m^c(C^-)}\inf_{B\in G_n^c(C^-)}\rho_\infty(A,B)=\sup_{A\in\overline{F_m^c(M)-F_m^c(C^-)}}\inf_{B\in\overline{G_n^c(M)-G_n^c(C^-)}}\rho_\infty(A,B).$$

Since we have

$$\sup_{A\in F_m^c(C^-)}\inf_{B\in G_n^c(C^-)}\rho_\infty(A,B)\geq\sup_{A\in\overline{F_m^c(M)-F_m^c(C^-)}}\inf_{B\in\overline{G_n^c(M)-G_n^c(C^-)}}\rho_\infty(A,B),$$

we only need to show that strict inequality does not hold, which will be done by contradiction.

Suppose the strict inequality holds, then there should exist $\varepsilon>0$, such that

$$\sup_{A\in F_m^c(C^-)}\inf_{B\in G_n^c(C^-)}\rho_\infty(A,B)>\sup_{A\in\overline{F_m^c(M)-F_m^c(C^-)}}\inf_{B\in\overline{G_n^c(M)-G_n^c(C^-)}}\rho_\infty(A,B)+\varepsilon.$$

Therefore, we can always find $A^*\in F_m^c(C^-)$ such that

$$\inf_{B\in G_n^c(C^-)}\rho_\infty(A^*,B)>\sup_{A\in F_m^c(C^-)}\inf_{B\in G_n^c(C^-)}\rho_\infty(A,B)-\varepsilon/2>\sup_{A\in\overline{F_m^c(M)-F_m^c(C^-)}}\inf_{B\in\overline{G_n^c(M)-G_n^c(C^-)}}\rho_\infty(A,B)+\varepsilon/2$$

By Lemma 4.40, we can also find $B^*\in G_n^c(C^-)$, such that $\rho_\infty(A^*,B^*)=\inf_{B\in G_n^c(C^-)}\rho_\infty(A^*,B)$. Thus,

$$\rho_\infty(A^*,B^*)>\sup_{A\in\overline{F_m^c(M)-F_m^c(C^-)}}\inf_{B\in\overline{G_n^c(M)-G_n^c(C^-)}}\rho_\infty(A,B)+\varepsilon/2. \tag{4.C.3}$$

Due to (4.C.1), we should always find $A^{**} \in \overline{F_m^c(M) - F_m^c(C^-)}$ and $B^{**} \in \overline{G_n^c(M) - G_n^c(C^-)}$, such that $\rho_\infty(A^{**}, B^{**}) = \rho_\infty(A^*, B^*)$. This contradicts (4.C.3). Therefore we have $\sup_{A\in F_m^c(M)} \inf_{B\in G_n^c(M)} \rho_\infty(A,B) = \sup_{A\in F_m^c} \inf_{B\in G_n^c} \rho_\infty(A,B)$ for $M > M_0$. Due to the symmetry of F_m^c and G_n^c, we also have $\sup_{A\in G_n^c(M)} \inf_{B\in F_m^c(M)} \rho_\infty(A,B) = \sup_{A\in G_n^c} \inf_{B\in F_m^c} \rho_\infty(A,B)$. Hence (4.2.10) follows. □

Proof of Lemma 4.4. According to Lemma 4.3, there exists $A^* \in F_m^c(M)$, $B^* \in G_n^c(M)$ such that $\rho_\infty(A^*, B^*) = \sup_{A\in F_m^c} \inf_{B\in G_n^c} \rho_\infty(A,B)$ for an sufficiently large $M > 0$, where $F_m^c(M)$ and $G_n^c(M)$ are defined in (4.C.2). By Lemma 4.40, we also have $\rho_\infty(A^*, B^*) = \inf_{A\in F_m^c} \rho_\infty(A, B^*)$. Thus, we have

$$\sup_{A\in F_m^c} \inf_{B\in G_n^c} \rho_\infty(A,B) \le \sup_{B\in G_n^c} \inf_{A\in F_m^c} \rho_\infty(A,B).$$

Similarly, by applying Lemmas 4.3 and 4.40, we can also show that

$$\sup_{A\in F_m^c} \inf_{B\in G_n^c} \rho_\infty(A,B) \ge \sup_{B\in G_n^c} \inf_{A\in F_m^c} \rho_\infty(A,B).$$

Thus, $\sup_{A\in F_m^c} \inf_{B\in G_n^c} \rho_\infty(A,B) = \sup_{B\in G_n^c} \inf_{A\in F_m^c} \rho_\infty(A,B)$. □

Let us note that the core of the proof relies on Lemma 4.40 and the boundedness $\mathcal{G}$. Therefore, Lemma 4.4 is generally true for the planar curves of two functions monotonic with respect to $\preceq$ with a compact region between them, which is important for us to prove Lemma 4.15.

The proof of the Theorem 4.5 relies on Lemma 4.40 and the expression of the Hausdorff distance given by Sendov and Beer (2012), i.e.

$$H(F_m^c, G_n^c) = \inf\{\epsilon : F_m^c \subset S(O, 2\epsilon) \oplus G_n^c, G_n^c \subset S(O, 2\epsilon) \oplus F_m^c\}.$$

This proof is similar to the proof of Lemma 3.8, with the additional consideration of Lemma 4.3 to ensure that there exists $A^* \in F_m^c$ and $B^* \in G_n^c$ such that the vector A^*B^* being parallel to OE_0 and $\rho_\infty(A^*, B^*) = d_0$ is obtainable, thus will be omitted.

Proof of Theorem 4.6. In order to prove (4.2.14), it suffices to show that for any $\varepsilon > 0$, when $\rho_\infty(X_i, \check{X}_i) < \varepsilon$ for $i = 1, \ldots, m$ and $\rho_\infty(Y_j, \check{Y}_j) < \varepsilon$ for $j = 1, \ldots, n$,

$$H(F_m^c, \check{F}_m^c) < 2\varepsilon, H(G_n^c, \check{G}_n^c) < 2\varepsilon. \tag{4.C.4}$$

This will be trivial since when $\rho_\infty(X_i,\check{X}_i)<\varepsilon$ for $i=1,\ldots,m$ and $\rho_\infty(Y_j,\check{Y}_j)<\varepsilon$ for $j=1,\ldots,n$, $\check{F}_m^c\subset F_m^c\oplus S(O,\varepsilon)$ and $\check{G}_n^c\subset G_n^c\oplus S(O,2\varepsilon)$. Hence by Theorem 4.5, (4.C.4) holds. □

Proof of Lemma 4.15. In order to show (4.2.21), according to the triangular inequality, it suffices to show that

$$H(F_m^c,G_n^c)=\max\{\inf_{B\in G_n^c}\rho_\infty(A_l,B),l=1,\ldots,2\upsilon\}$$

when $\mathcal{G}$ is compact. This statememt may seems not to be direct. Alternatively, for an arbitrary $\varepsilon>0$, we can always find a continuous $\hat{G}\in C(\mathbb{R}^k)$ being strictly monotonic with respect to $\preceq$, such that its planar curve $\hat{G}^c\subset G_n^c\oplus S(O,\varepsilon)$ and the region between F_m^c and $\hat{G}^c$, $\mathcal{G}(F_m^c,\hat{G})$, is compact. Since $\hat{G}^c\subset G_n^c\oplus S(O,\varepsilon)$, we should have

$$\begin{aligned}&|H(F_m^c,G_n^c)-H(F_m^c,\hat{G}^c)|<\varepsilon,\\ &\qquad\text{and }\left|\inf_{B\in\hat{G}^c}\rho_\infty(A_l,B)-\inf_{B\in G_n^c}\rho_\infty(A_l,B)\right|<\varepsilon,\text{ for }l=1,\ldots,2\upsilon.\end{aligned}\tag{4.C.5}$$

Since $\hat{G}$ is monotonic with respect to $\preceq$ and $\mathcal{G}(F_m^c,\hat{G})$ is compact, by Lemma 4.4, we have

$$H(F_m^c,\hat{G}^c)=\sup_{A\in F_m^c}\inf_{B\in\hat{G}}\rho_\infty(A,B).\tag{4.C.6}$$

By (4.C.5), (4.C.6) and the arbitrariness of ε, it remains to show that

$$\sup_{A\in F_m^c}\inf_{B\in\hat{G}}\rho_\infty(A,B)=\max\{\inf_{B\in\hat{G}}\rho_\infty(A_l,B),l=1,\ldots,2\upsilon\}.\tag{4.C.7}$$

To show this, we first consider $A\in F_m^c$ that is locally flat, i.e. $A_0=(x_{A_0},F_m(x_{A_0}))^T$, $x_{A_0}\in\mathbb{R}^k$ and there exists $\delta>0$ such that $\forall\, x\in\{x:\rho_\infty^*(x,x_{A_0})<\delta\}$, $F_m(x)=F_m(x_{A_0})$, where ρ_∞^* is ρ_∞ restricted on $\mathbb{R}^k$. Therefore, we can always find $A_1=(x_{A_1},z_{A_1})^T,A_2=(x_{A_2},z_{A_2})^T\in F_m^c$ such that $x_{A_1}\prec x_{A_0}\prec x_{A_2}$ and $z_{A_0}=z_{A_1}=z_{A_2}$. Since $\hat{G}$ is strictly monotonic with respect to $\preceq$, we should have either

$$\inf_{B\in\hat{G}^c}\rho_\infty(A_0,B)>\inf_{B\in\hat{G}^{c\prime}}\rho_\infty(A_1,B)\text{ or }\inf_{B\in\hat{G}^c}\rho_\infty(A_0,B)>\inf_{B\in\hat{G}^{c\prime}}\rho_\infty(A_2,B)$$

to be true. Therefore the supremum cannot be achieved at A_0. Similarly, for A_0' at

the vertical part of F_m^c, if A_0' is at the interior, we can also find A_1' and A_2', such that $x_{A_0'} = x_{A_1'} = x_{A_2'}$ but $z_{A_1'} < z_{A_0'} < z_{A_2'}$, once again by the strict monotonic of $\hat{G}$, either

$$\inf_{B\in\hat{G}^c} \rho_\infty(A_0', B) > \inf_{B\in\hat{G}^{c\prime}} \rho_\infty(A_1', B) \text{ or } \inf_{B\in\hat{G}^c} \rho_\infty(A_0', B) > \inf_{B\in\hat{G}^{c\prime}} \rho_\infty(A_2', B)$$

should be true. Therefore, the supremum cannot be achieved at A_0. By using the same reasoning, the supremum is neither achieved at saddle points. However, since $\mathcal{G}(F_m^c, \hat{G})$ is compact, the supremum is achievable. Therefore, the supremum can only be achieved at A_l, $l \in 1, \ldots, 2\upsilon$. Hence, (4.C.7) should hold. □

Proof of Theorem 4.19. Consider the vertices $A_1, A_2, \ldots, A_{2\upsilon}$ and $B_1, B_2, \ldots, B_{2\nu}$ of the curves F_m^c and G_n^c respectively. Let us fix an arbitrary vertex A_{2k}, $k = 0, 1, 2, \ldots, \upsilon$.

Let us take the straight line with points $\{(x, z) : z - \frac{k}{m} = x_{k,m} - x\}$, passing through the vertex A_{2k}, with coordinates $(x_{k,m}, \frac{k}{m})$, where $x_{k,m}$ is the realization of the k-th order statistic, $X_{(k)}$ of the random sample $\boldsymbol{X}_m$.

It is easy to see that the intersection of the sets G_n^c and $\{(x, z) : z - \frac{k}{m} = x_{k,m} - x\}$ consists of only one point which we denote by P_{2k}, i.e. $P_{2k} = G_n^c \cap \{(x, z) : z - \frac{k}{m} = x_{k,m} - x\}$.

Let us consider the two cases:

1. P_{2k} on the vertical line $\{(x_{k,m}, z) : z \in \mathbb{R}\}$. In this case it is easy to see that

$$\begin{aligned} \rho_\infty(A_{2k}, P_{2k}) &\leq |F_m(x_{k,m}) - G_n(x_{k,m})| \\ &= |\frac{k}{m} - G_n(x_{k,m})| \end{aligned} \tag{4.C.8}$$

From (4.C.8), we obtain that

$$\rho_\infty(A_{2k}, P_{2k}) \leq \sup_{1\leq k\leq n} |\frac{k}{m} - G_n(x_{k,m})|, \tag{4.C.9}$$

for $k = 1, \ldots, m$. Therefore, we have from (4.C.9) that

$$\sup_{1\leq k\leq m} \rho_\infty(A_{2k}, P_{2k}) \leq \sup_{1\leq k\leq m} |\frac{k}{m} - G_n(x_{k,m})| \tag{4.C.10}$$

Next, we provide an auxiliary lemma that we apply to the left-hand side of (4.C.10).

Lemma 4.41. *We have*

$$\sup_{1\le k\le m} \rho_\infty(A_{2k}, P_{2k}) = H(F_m^c, G_n^c)$$

Now, from Lemma 4.41 and equality (4.C.10), we can see that

$$H(F_m^c, G_n^c) \le \sup_{1\le k\le m} |\frac{k}{m} - G_n(x_{k,m})| = \sup_{-\infty<x<\infty} |F_m(x) - G_n(x)| = D_{m,n}$$

2. Let us now consider the second case, i.e. P_{2k} on the horizontal line $\{(x, \frac{l}{n})\}$. On the other hand, $P_{2k} \in z - \frac{k}{m} = x_{k,m} - x$. Therefore, $\frac{l}{n} - \frac{k}{m} = x_{k,m} - x \Rightarrow x = x_{k,m} + \frac{k}{m} - \frac{l}{n}$, i.e. $P_{2k} = (x_{k,m} + \frac{k}{m} - \frac{l}{n}, \frac{l}{n})$.

   Let us now evaluate $H(A_{2k}, P_{2k})$. We have $h(A_{2k}, P_{2k}) = \max(|x_{k,m} + \frac{k}{m} - \frac{l}{n} - x_{k,m}|, |\frac{l}{n} - \frac{k}{m}|) \equiv |\frac{k}{m} - \frac{l}{n}| = |F_m(x_{k,m}) - G_n(x_{k,m})|$. Therefore,

$$\begin{aligned}\sup_{1\le k\le m} \rho_\infty(A_{2k}, P_{2k}) &= \sup_{1\le k\le m} |F_m(x_{k,m}) - G_n(x_{k,m})| \\ &= \sup_{-\infty<x<+\infty} |F_m(x) - G_n(x)| = D_{m,n},\end{aligned}$$

   which completes the proof of inequality (4.2.25).

To show that (4.2.26) holds, recall that from Lemma 4.4, we have

$$\begin{aligned}H(F_m^c, G_n^c) &= \max\left[\sup_{A\in F_m^c}\inf_{B\in G_n^c} \rho_\infty(A,B), \sup_{B\in G_n^c}\inf_{A\in F_m^c} \rho_\infty(A,B)\right] \\ &= \sup_{A\in F_m^c}\inf_{B\in G_n^c} \rho_\infty(A,B) = \sup_{B\in G_n^c}\inf_{A\in F_m^c} \rho_\infty(A,B)\end{aligned}$$

Equality (4.2.26) follows noting that, if condition (4.2.27) holds, we have

$$\rho_\infty(A,B) = \max\{|x_A - x_B|, |y_A - y_B|\} = |y_A - y_B|$$

Therefore $H(F_m^c, G_n^c) = \sup_{-\infty<x<+\infty} |F_m(x) - G_n(x)|$ □

Proof of Lemma 4.23. Let us note,

$$H\left(\tilde{A}_{2l},\tilde{E}_{2l}\right)=\rho_{\infty}(\tilde{A}_{2l},\tilde{E}_{2l}), \tag{4.C.11}$$

the geometric interpretation of the latter distance, $\rho_{\infty}\left(\tilde{A}_{2l},\tilde{E}_{2l}\right)$, is the side of the square. In order to prove (4.3.6), we will consider the two possible ways in which the curves $\tilde{F}_m^c$ and $\tilde{G}_n^c$ may be positioned with respect to each other, which corresponds to either i being odd, i.e. the segment $\left[\tilde{B}_{i-1},\tilde{B}_i\right]$, crossed by the line $\mathcal{L}_l$, being horizontal, or i being even, i.e., $\left[\tilde{B}_{i-1},\tilde{B}_i\right]$ being vertical.

Case 1, $i=2p+1$ odd, i.e., $\left[\tilde{B}_{i-1},\tilde{B}_i\right]$ horizontal. The case is illustrated in

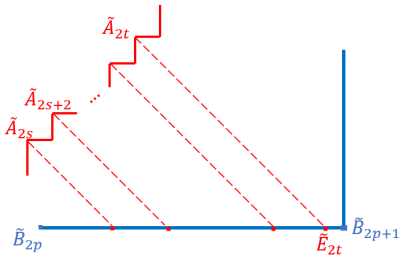


Figure 4.11: Graphical illustration of Case 1, $i=2p+1$ odd, i.e., $[\tilde{B}_{i-1},\tilde{B}_i]$ horizontal

Figure 4.11, from where one can see that the maximum, in (4.3.6) is achieved in the last vertex, $\tilde{A}_{2t}$ of the batch $\bar{A}_i$. This holds, because, by contruction of the curve $\tilde{F}_m^c$, its consecutive vertices $\tilde{A}_{2j}$, $j=s,s+1,\ldots,t$, deviate further away from the horizonal segment $[\tilde{B}_{i-1},\tilde{B}_i]$ of the curve $\tilde{G}_n^c$, as j increase to t. Setting $j=t$ in (4.C.11), the projection $\lambda_{\tilde{A}_{2t}}$, on the Ox axis, of the vertex, $\tilde{A}_{2t}(\sum_{k=1}^t\tilde{a}_k,\sum_{k=1}^t\tilde{b}_k)^T$, is $\lambda_{\tilde{A}_{2t}}=\sum_{k=1}^t(\tilde{a}_k+\tilde{b}_k)$.

From the above considerations, we can write that

$$\begin{aligned}&\max_{A\in\bar{A}_i}\inf_{B\in\tilde{G}_n^c}\rho_{\infty}(A,B)=\max_{s\le j\le t}\inf_{B\in\tilde{G}_n^c}\rho_{\infty}(\tilde{A}_{2j},B)\\ =&\rho_{\infty}(\tilde{A}_{2t},\tilde{E}_{2t})=y_{\tilde{A}_{2t}}-y_{\tilde{E}_{2t}}\\ =&\sum_{k=1}^t\tilde{b}_k-y_{\tilde{B}_{2p}}=\sum_{k=1}^t\tilde{b}_k-\sum_{k=1}^p\tilde{d}_k\end{aligned}$$

Case 2, $i=2(s+1)$ even, i.e., $[B_{i-1},B_i]$ vertical. This case is illustrated in

Figure 4.12, from where one can see that the maximum in (4.3.6) is achieved in the first vertex, $\tilde{A}_{2s}$ of the batch $\bar{A}_i$. This holds, because, by construction of the

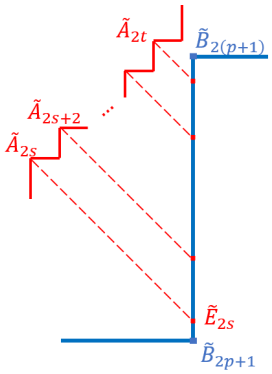


Figure 4.12: Graphical illustration of Case 2, $i = 2(p+1)$ even, i.e., $[\tilde{B}_{i-1}, \tilde{B}_i]$ vertical

curve $\tilde{F}^c_m$, its consecutive vertices, $\tilde{A}_{2j}$, $j = s, s+1, \ldots, t$, are positioned closer to the vertical segment $[\tilde{B}_{i-1}, \tilde{B}_i]$ of the curve $\tilde{G}^c_n$, as j increases to t. Setting $j = s$ in (4.C.11), the projection, $\lambda_{\tilde{A}_{2s}}$ on the Ox axis, of the vertex, $\tilde{A}_{2s}(\sum_{k=1}^{s} \tilde{a}_k, \sum_{k=1}^{s} \tilde{b}_k)^T$, is $\lambda_{\tilde{A}_{2s}} = \sum_{k=1}^{s}(\tilde{a}_k + \tilde{b}_k)$.

Following from the above considerations, we can write that

$$\begin{aligned}
&\max_{A \in \tilde{A}_i} \inf_{B \in \tilde{G}^c_n} \rho_\infty(A, B) = \max_{s \le j \le t} \inf_{B \in \tilde{G}^c_n} \rho_\infty(\tilde{A}_{2j}, B) \\
=&\rho_\infty(\tilde{A}_{2s}, \tilde{E}_{2s}) = x_{\tilde{A}_{2s}} - x_{\tilde{E}_{2s}} \\
=&x_{\tilde{B}_{2(p+1)}} - \sum_{k=1}^{s} \tilde{a}_k = a_1 + \sum_{k=1}^{p+1} \tilde{c}_k - \sum_{k=1}^{s} \tilde{a}_k = \sum_{k=1}^{p+1} \tilde{c}_k - \sum_{k=2}^{s} \tilde{a}_k
\end{aligned}$$

This completes the proof of the lemma. □

The proof of Theorem 4.24 will be given after the proof of Theorem 4.28.

Lemma 4.42. *When $\boldsymbol{X}_m$ comes from distribution F and $\boldsymbol{Y}_n$ comes from distribution G, we have*

$$\sup_{t \in \mathbb{R}} \left| \sqrt{\frac{mn}{m+n}} (E_{m+n}(t) - E(t)) - \left(\sqrt{1-\eta}\mathbb{B}_0(F(t)) + \sqrt{\eta}\mathbb{B}'_0(G(t)) \right) \right| \to 0 \ a.s. \tag{4.C.12}$$

as $m+n \to \infty$ and $\frac{m}{m+n} \to \eta$, under norm l^∞, where $\mathbb{B}_0(F)$ and $\mathbb{B}'_0(G)$ are independent Brownian bridges. Furthermore, when $\boldsymbol{R}_{m+n}$ is uniformly distributed, we also have

$$\mathbb{P}\left\{ \left[\sqrt{\frac{mn}{m+n}} \left(G^\dagger_n(t) - E_{m+n}(t) \right) \right) \middle| \boldsymbol{Z}_{m+n} \right] \xrightarrow{w} (1-\eta)\mathbb{B}_0(E(t)) \right\} = 1 \tag{4.C.13}$$

where the weak convergence $\xrightarrow{w}$ *is considered over the indicator functions of intervals of* $\mathbb{R}$*, i.e.* $\{\mathbb{1}_{(x,y]}(\cdot) : -\infty \le x < y \le +\infty\}$.

Proof. The proof of (4.C.12) is a natural extension of Theorem 3 and Corollary of Komlós et al. (1975), thus will be omitted. Since the class $\{\mathbb{1}_{(x,y]}(\cdot) : -\infty \le x < y \le +\infty\}$ has a square integrable envelope function $\mathbb{1}_{\mathbb{R}}$ with respect to measure μ_F and μ_G, i.e. $\int \mathbb{1}_{\mathbb{R}}^2 d\mu_F = \int \mathbb{1}_{\mathbb{R}}^2 d\mu_G = 1 < \infty$, where F and G are the probability measure of the distribution F and G, (4.C.13) follows by applying Theorem 3.8.2 in van der Vaart and Wellner (2023). □

Proof of Theorem 4.28. Since metric H coincides with the Levy metric, according to the definition of the Levy metric (see in Remark 4.22)

$$\mathcal{H}_{m,n}^{\dagger} = \inf_{\varepsilon}\Big\{\varepsilon : F_m^{\dagger}(t-\varepsilon) - \varepsilon \le G_n^{\dagger}(t) \le F_m^{\dagger}(t+\varepsilon) + \varepsilon, \text{ for all } t\Big\}$$

Hence,

$$\begin{aligned}\sqrt{\frac{mn}{m+n}}\mathcal{H}_{m,n}^{\dagger} = \inf_{\varepsilon}\Bigg\{\varepsilon : -\varepsilon \le \sqrt{\frac{mn}{m+n}}\Big[G_n^{\dagger}(t) - F_m^{\dagger}\Big(t - \varepsilon/\sqrt{\frac{mn}{m+n}}\Big)\Big] \\ \text{and } \sqrt{\frac{mn}{m+n}}\Big[G_n^{\dagger}(t) - F_m^{\dagger}\Big(t + \varepsilon/\sqrt{\frac{mn}{m+n}}\Big)\Big] \le \varepsilon, \text{ for all } t\Bigg\}\end{aligned}$$

Thus

$$\begin{aligned}\mathbb{P}\Big\{\sqrt{\frac{mn}{m+n}}\mathcal{H}_{m,n}^{\dagger} > x \Big| \boldsymbol{Z}_{m+n}\Big\} = \mathbb{P}\Bigg\{\forall\, \varepsilon < x, \exists\, t_0 \in \mathbb{R}, \text{such that} \\ \sqrt{\frac{mn}{m+n}}\Big[G_n^{\dagger}(t_0) - F_m^{\dagger}\Big(t_0 - \varepsilon/\sqrt{\frac{mn}{m+n}}\Big)\Big] < -\varepsilon \text{ or} \\ \sqrt{\frac{mn}{m+n}}\Big[G_n^{\dagger}(t_0) - F_m^{\dagger}\Big(t_0 + \varepsilon\sqrt{\frac{mn}{m+n}}\Big)\Big] > \varepsilon \Big| \boldsymbol{Z}_{m+n}\Bigg\}\end{aligned} \tag{4.C.14}$$

Note that $F_m^\dagger(t) - E_{m+n}(t) = \frac{n}{m+n}(F_m^\dagger(t) - G_n^\dagger(t))$. Then

$$\begin{aligned}
&\sqrt{\frac{mn}{m+n}}\left[G_n^\dagger(t) - F_m^\dagger\left(t-\varepsilon/\sqrt{\frac{mn}{m+n}}\right)\right] \\
=&\sqrt{\frac{mn}{m+n}}\Bigg[(G_n^\dagger(t) - E_{m+n}(t)) + \left(E_{m+n}\left(t-\varepsilon/\sqrt{\frac{mn}{m+n}}\right) - F_m^\dagger\left(t-\varepsilon/\sqrt{\frac{mn}{m+n}}\right)\right) \\
&+ (E_{m+n}(t) - E(t)) + \left(E\left(t-\varepsilon/\sqrt{\frac{mn}{m+n}}\right) - E_{m+n}\left(t-\varepsilon/\sqrt{\frac{mn}{m+n}}\right)\right) \\
&+ E(t) - E\left(t-\varepsilon/\sqrt{\frac{mn}{m+n}}\right)\Bigg]
\end{aligned}$$

For simplicity, let $X(t,h) = \int \mathbb{1}_{[t,t+h)} d[\sqrt{\eta}\mathbb{B}_0(F) + \sqrt{1-\eta}\mathbb{B}_0(G)]$, therefore $\sup_{t\in\mathbb{R}} X(t,h)$ is the modulus of continuity of the Brownian Bridge $\sqrt{\eta}\mathbb{B}_0(F) + \sqrt{1-\eta}\mathbb{B}_0(G)$. Since both F and G are uniformly continuous, by Lévy's modulus of continuity theorem, we have $\sup_{t\in\mathbb{R}} X(t,h) \to 0$ almost surely as $h \to 0$. According to (4.C.12) in Lemma 4.42, we have

$$\sqrt{\frac{mn}{m+n}}\left(E_{m+n}(t) - E(t) + E\left(t-\varepsilon/\sqrt{\frac{mn}{m+n}}\right) - E_{m+n}\left(t-\varepsilon/\sqrt{\frac{mn}{m+n}}\right)\right) \to 0 \ a.s.$$

In addition, since F and G have bounded densities, we have

$$\sup_{t\in\mathbb{R}} \left|\sqrt{\frac{mn}{m+n}}(E(t) - E(t-\varepsilon/\sqrt{\frac{mn}{m+n}})) - \varepsilon\delta^-(t)\right| \to 0.$$

Thus given the pooled sample $\boldsymbol{Z}_{m,n}$,

$$\sup_{t\in\mathbb{R}} \left|\sqrt{\frac{mn}{m+n}}\left(E_{m+n}(t) - E_{m+n}\left(t-\varepsilon/\sqrt{\frac{mn}{m+n}}\right)\right) - \varepsilon\delta^-(t)\right| \to 0, \tag{4.C.15}$$

almost surely. Additionally, by the continuity of E and (4.C.13), we have

$$\begin{aligned}
\mathbb{P}\Bigg\{\Bigg[\sqrt{\frac{mn}{m+n}}\Bigg(&G_n^\dagger(t) - E_{m+n}(t) + \\
&E_{m+n}\left(t-\varepsilon/\sqrt{\frac{mn}{m+n}}\right) - F_m^\dagger\left(t-\varepsilon/\sqrt{\frac{mn}{m+n}}\right)\Bigg)\Bigg|\boldsymbol{Z}_{m+n}\Bigg] \xrightarrow{w} \mathbb{B}_0(E(t))\Bigg\} = 1
\end{aligned} \tag{4.C.16}$$

By arranging (4.C.14), (4.C.15) and (4.C.16), we obtain

$$
\begin{aligned}
&\mathbb{P}\left\{\sqrt{\frac{mn}{m+n}}\mathcal{H}_{m,n}^{\dagger} > x \middle| \boldsymbol{Z}_{m+n}\right\} \\
\to&\mathbb{P}\{\forall\, \varepsilon < x, \exists\, t,\ -\varepsilon > \mathbb{B}_0 \circ E + \varepsilon e(t),\ \text{or } \mathbb{B}_0 \circ E - \varepsilon e(t) > \varepsilon\} \\
=&\mathbb{P}\left\{\exists\, t_0 \in [0,1], \forall\, \varepsilon < x,\ -\varepsilon - \varepsilon e\left(E^{-1}(t_0)\right) > \mathbb{B}_0(t_0) \text{ or } \mathbb{B}_0(t_0) > \varepsilon + \varepsilon e\left(E^{-1}(t_0)\right)\right\} \\
=&\mathbb{P}\left\{\exists\, t_0 \in [0,1], \inf_{\varepsilon<x}\left[-\varepsilon - \varepsilon e\left(E^{-1}(t_0)\right)\right] > \mathbb{B}_0(t_0) \text{ or } \mathbb{B}_0(t_0) > \sup_{\varepsilon<x}\left[\varepsilon + \varepsilon e\left(E^{-1}(t_0)\right)\right]\right\} \\
=&\mathbb{P}\left\{\exists\, t_0 \in [0,1], -x(1+e(E^{-1}(t_0))) > \mathbb{B}_0(t_0) \text{ or } \mathbb{B}_0(t_0) > x[1+e(E^{-1}(t_0))]\right\}
\end{aligned}
$$

Therefore, we have

$$
\begin{aligned}
&\lim_{m,n\to\infty, \frac{m}{m+n}\to\eta} \mathbb{P}\left\{\sqrt{\frac{mn}{m+n}}\mathcal{H}_{m,n}^{\dagger} > x\right\} \\
=&\mathbb{P}\left\{\exists\, t_0 \in [0,1], -x\left[1+e(E^{-1}(t_0))\right] > \mathbb{B}_0(t_0) \text{ or } \mathbb{B}_0(t_0) > x\left[1+e(E^{-1}(t_0))\right]\right\} \\
=&1 - \mathbb{P}\left\{-x[\left[1+e(E^{-1}(t_0))\right] \le \mathbb{B}(t) \le x\left[1+e(E^{-1}(t_0))\right],\ \forall\, 0 \le t \le 1\right\}.
\end{aligned}
$$

This finishes the proof of the required theorem. □

Proof of Theorem 4.24. In order to show (4.3.11) holds under the condition of Theorem 4.24, according to the Portmanteau Theorem (cf. Theorem 1.3.4 in van der Vaart and Wellner, 2023), one only needs to prove that

$$
\mathbb{P}\left(\sqrt{\frac{mn}{m+n}}\mathcal{H}_{m,n} \le q\right) \to \mathbb{P}\left\{|\mathbb{B}(t)| \le x\left[1+e(E^{-1}(t))\right],\ \forall\, 0 \le t \le 1\right\} \tag{4.C.17}
$$

when $\boldsymbol{X}_m$ and $\boldsymbol{Y}_n$ come from $\eta F(x) + (1-\eta)G(x)$. Note that when the pooled sample $\boldsymbol{Z}_{m+n} = \{X_1, \ldots, X_m, Y_1, \ldots, Y_n\}$ come from the pooled distribution E,

$$
\sqrt{\frac{mn}{m+n}}(E_{m+n} - E) \xrightarrow{w} \mathbb{B}_0(E) \tag{4.C.18}
$$

under norm l^{∞}. Therefore, we have

$$
\sqrt{\frac{mn}{m+n}}\left(G_n(t) - E_{m+n}(t)\right) \xrightarrow{w} (1-\eta)\mathbb{B}_0(E(t)) \tag{4.C.19}
$$

over the class $\{\mathbb{1}_{(x,y]}(\cdot) : -\infty \le x < y \le +\infty\}$, which is equivelent to (4.C.16) with respect to respect to F_m and G_n. Therefore, the required convergence (4.C.17) is

obtained by arranging (4.C.14), (4.C.15) and (4.C.16) with respect to F_m, G_n. □

Proof of Theorem 4.25. Under the contiguous alternative, since $\rho_{C^1}(F,G) = o(1)$, we have $\sup_x |f(x) - e(x)| \to 0$. We can show that (4.C.19) still holds (although (4.C.18) no longer holds).

For a fixed x, let $b = \mathbb{P}\{|\mathbb{B}(t)| \le x[1 + f(F^{-1}(t))], \forall\, 0 \le t \le 1\}$ and let sequences

$$a_{m,n,k} = \mathbb{P}\left\{\sqrt{\frac{mn}{m+n}}\mathcal{H}^{\dagger}_{m,n} \le x \middle| \boldsymbol{Z}_{m+n}\right\},$$

when $\boldsymbol{X}_m$ and $\boldsymbol{Y_n}$ correspondingly come from F and $G(;k)$,

$$b_k = \mathbb{P}\left\{|\mathbb{B}(t)| \le x\left[1 + e_k(E_k^{-1}(t))\right], \forall\, 0 \le t \le 1\right\},$$

where $E_k(x) = \eta F(x) + (1-\eta)G(x;k)$ with density e_k. Therefore, to show (4.3.13), it suffices to show that $a_{m,n,n} \to b$ as $m, n \to \infty$. Note that $\rho_{C^1}(F, G(;k)) \to 0$, $\rho_{C^1}(F, E_k) \to 0$, by the property of Brownian bridge, we have $b_k \to b$. By Theorem 4.28, we also have that $a_{m,n,k} \to b_k$ as $m, n \to \infty$ and $\frac{m}{m+n} \to \eta$. Therefore, it remains to show that $a_{m,n,k} \to b_k$ uniformly as $m, n \to \infty$ and $\frac{m}{m+n} \to \eta$, or equivalently that the convergence in (4.C.13) is uniform.

To show (4.C.13) is uniform convergence, we need to show that (i) the speed of convergence in (4.C.13) does not depend on k for a finite subset $\mathcal{F}_0 \subset \mathcal{F}$ and (ii) the left-hand side process of (4.C.13) is stochastic equicontinuity over the class $\mathcal{F} = \{\mathbb{1}_{(x,y]}(\cdot) : -\infty \le x \le y \le +\infty\}$ and uniformly with respect to k.

In fact, (i) is ensured by the existence of a square-integrable envelope function $\mathbb{1}_{\mathbb{R}}(\cdot)$ of $\mathcal{F}$. To show (ii), we further denote $\mathcal{F}_{\delta,k} = \{f - g : f, g \in \mathcal{F}, \int (f-g)^2 dE_k \le \delta\}$ and $\mathcal{F}_\delta = \{f - g : f, g \in \mathcal{F}, \int (f-g)^2 dE \le \delta\}$. Since $\rho_{C^1}(F, G(;k)) \to 0$, for any $\varepsilon > 0$, there exists K, such that for any $k > K$, $\mathcal{F}_{\delta,k} \subset \mathcal{F}_{(1+\varepsilon)\delta}$. Note that $\mathcal{F}$ has an envelope function $\mathbb{1}_{\mathbb{R}}$ that is square-intergrable with respect to each measure $\mu_{G(;k)}$ and μ_F, by the inequalities established on page 508 van der Vaart and Wellner (2023), we have

$$\begin{aligned}\mathbb{E}\left[\sup_{f\in\mathcal{F}_\delta}\int f d(\mu_{\boldsymbol{X}^{\dagger}_m} - \mu_{\boldsymbol{Z}_{m+n}}) \middle| \boldsymbol{Z}_{m+n}\right] \le \\ C_0\left(\mathbb{E}\left[\sup_{f\in\mathcal{F}_\delta}\int f dZ_F\right] + \sqrt{(1-\eta)/\eta}\mathbb{E}\left[\sup_{f\in\mathcal{F}_\delta}\int f dZ_{G(;k)}\right]\right),\end{aligned} \tag{4.C.20}$$

for almost every $\boldsymbol{Z}_{m+n}$ where Z_F and $Z_{G(;k)}$ are tight Brownian processes, $\mu_{\boldsymbol{X}_m^\dagger}$ is the empirical measure corresponding to $\boldsymbol{X}_m^\dagger$, and C_0 is a universal constant that does not depend on the measures F and G. Based on that $\rho_{C^1}(E, E_k) \to 0$ and that $Z_{G(;k)}$ have continuous sample paths with respect to the semi-metric $L^2(E_k)$, one can easily show that both Z_F and $Z_{G(;k)}$ have uniformly continuous sample paths with respect to the semi-metric $L^2(E)$. Therefore, the RHS of (4.C.20) converges to 0 when $\delta \downarrow 0$. Therefore $\mathbb{E}\left[\sup_{f\in\mathcal{F}_\delta} \int f d(\mu_{\boldsymbol{X}_m^\dagger} - \mu_{\boldsymbol{Z}_{m+n}}) \middle| \boldsymbol{Z}_{m+n}\right] \to 0$ as $\delta \downarrow 0$. In addition, according to the fact that $\mathcal{F}_{\delta,k} \subset \mathcal{F}_{(1+\varepsilon)\delta}$ holds for sufficiently large k, one can show that

$$\lim_{\delta\downarrow 0} \sup_{k,m,n} \left\{ \mathbb{E}\left[\sup_{f\in\mathcal{F}_{\delta,k}} \int f d(\mu_{\boldsymbol{X}_m^\dagger} - \mu_{\boldsymbol{Z}_{m+n}}) \middle| \boldsymbol{Z}_{m+n} \right] \right\} = 0.$$

Therefore (ii) holds.

Then by Theorem 1.5.4 and Lemma 1.5.9 of van der Vaart and Wellner (2023), one can show that the weak convergence in (4.C.13) holds over $\mathcal{F}$ when (i) finite convergence and (ii) stochastic equicontinuity satisfy. In addition, we have shown that both the finite convergence and stochastic equicontinuity do not depend on k. Therefore, we can show that $a_{m,n,k} \to b_k$ uniformly in k. Hence, the theorem follows. □

Lemma 4.43. *For a fixed x, let $\boldsymbol{X}_m$ come from F and let $\boldsymbol{Y}_n$ come from a contigous alternative $G(x;n) = F(x) + \frac{1}{\sqrt{n}}\delta(x)$, $\delta \in C(\mathbb{R})$, we have*

$$\mathbb{P}\left(\sqrt{\frac{mn}{m+n}}\mathcal{H}_{m,n} > x\right) \to 1 - \mathbb{P}\{|\mathbb{B}(t) + \sqrt{\eta}\delta(t)| \le x[1 + e(E^{-1}(t))],\ \forall\, 0 \le t \le 1\} \tag{4.C.21}$$

Proof. We have

$$\begin{aligned}
&\sqrt{\frac{mn}{m+n}}\left[G_n(t) - F_m\left(t - \varepsilon/\sqrt{\frac{mn}{m+n}}\right)\right] \\
=&\sqrt{\frac{mn}{m+n}}\left[(G_n(t) - G(t;n)) + \left(F\left(t - \varepsilon/\sqrt{\frac{mn}{m+n}}\right) - F_m\left(t - \varepsilon/\sqrt{\frac{mn}{m+n}}\right)\right)\right. \\
&\left. + G(t;n) - F\left(t - \varepsilon/\sqrt{\frac{mn}{m+n}}\right)\right]
\end{aligned}$$

We can show that

$$\sqrt{\frac{mn}{m+n}}\left[(G_n(t)-G(t;n))+\left(F\left(t-\varepsilon/\sqrt{\frac{mn}{m+n}}\right)-F_m\left(t-\varepsilon/\sqrt{\frac{mn}{m+n}}\right)\right)\right]\xrightarrow{w}\mathbb{B}_0(F) \tag{4.C.22}$$

almost surely. In addition, we have

$$\sqrt{\frac{mn}{m+n}}\left[G(t;n)-F(t-\varepsilon/\sqrt{\frac{mn}{m+n}})\right]\to\sqrt{\eta}\delta(t)+\varepsilon f(t) \tag{4.C.23}$$

According to (4.C.12) in Lemma 4.42 and that $\delta\in C(\mathbb{R})$, we can also show that the convergence in (4.C.22) and (4.C.23) are uniform convergence. Therefore, (4.C.21) follows. □

Proof of Theorem 4.26. For a fixed null, by (4.3.11) and definitions of q^*, it is not difficult to show that (4.3.12) holds. Therefore, we only need to prove (4.3.13).

For a fixed p and fixed alternative, it is easy to show that both $q^\dagger$ and q^* are bounded from above, and for any $\varepsilon>0$, $\mathbb{P}(\sqrt{\frac{mn}{m+n}}\mathcal{H}_{m,n}>\varepsilon)\to 1$, therefore (4.3.13) follows.

Under a contiguous alternative $\rho_{C^1}(F,G(\cdot;n))=O(n^{-1/2})$, (4.3.13) follows by applying Lemma 4.43 and Theorem 4.25. □

Proof of Theorem 4.32. For simplicity, denote by

$$A_m=\frac{\psi_1-\psi_2}{Q_{\boldsymbol{X}_m^\dagger}(\psi_1)-Q_{\boldsymbol{X}_m^\dagger}(\psi_2)}\text{ and }B_n=\frac{\psi_1-\psi_2}{Q_{\boldsymbol{Y}_n^\dagger}(\psi_1)-Q_{\boldsymbol{Y}_n^\dagger}(\psi_2)}.$$

Given $\boldsymbol{Z}_{m+n}$, the number of observations $N_{\boldsymbol{X}_m^\dagger}(\psi)$ in $\boldsymbol{X}_m^\dagger$ that are less than $E_{m+n}^{-1}(\psi)$ follows a hypergeometric distribution, i.e. $N_{\boldsymbol{X}_m^\dagger}(\psi)=\text{Hypergeometric}(m+n,\lfloor m\psi\rfloor,m)$. By the law of large numbers, $F_m(E_{m+n}^{-1}(\psi))=N_{\boldsymbol{X}_m^\dagger}(\psi)/m\xrightarrow{\mathbb{P}}\psi$. Therefore given $\boldsymbol{Z}_{m+n}$,

$$|Q_{X_m}^\dagger(\psi)-E_{m+n}^{-1}(\psi)|\le C(|N_{\boldsymbol{X}_m^\dagger}(\psi)/m-\psi|+\frac{1}{m+n})\xrightarrow{\mathbb{P}}0.$$

Hence, we have $Q_{X_m}^\dagger(\psi)\xrightarrow{\mathbb{P}}E_{m+n}^{-1}(\psi)$, which then imply $Q_{\boldsymbol{X}_m^\dagger}(\psi_1)-Q_{\boldsymbol{X}_m^\dagger}(\psi_2)\xrightarrow{\mathbb{P}}E_{m+n}^{-1}(\psi_1)-E_{m+n}^{-1}(\psi_2)$. By the continuous mapping theorem, given $\boldsymbol{Z}_{m+n}$, we have $A_m-\frac{\psi_1-\psi_2}{E_{m+n}^{-1}(\psi_1)-E_{m+n}^{-1}(\psi_2)}\xrightarrow{\mathbb{P}}0$ The same result holds with respect to B_n. Thus, given

$\boldsymbol{Z}_{m+n}$,

$$\max(A_m, B_n) - \frac{\psi_1 - \psi_2}{E_{m+n}^{-1}(\psi_1) - E_{m+n}^{-1}(\psi_2)} \xrightarrow{\mathbb{P}} 0 \tag{4.C.24}$$

By the Glivenko-Cantelli theorem, $\sup_t |F_m(t) - F(t)| \to 0$ and $\sup_t |G_n(t) - G(t)| \to 0$ almost surely. Therefore, it is clear that $\sup_t |E_{m+n}(t) - E(t)| \to 0$ almost surely. Hence, for an arbitrary $\psi \in (0,1)$, one have $E_{m+n}^{-1}(\psi) \to E^{-1}(\psi)$ almost surely. Therefore by continuous mapping theorem, $\frac{\psi_1 - \psi_2}{E_{m+n}^{-1}(\psi_1) - E_{m+n}^{-1}(\psi_2)} \xrightarrow{\mathbb{P}} \sigma_0$. Thus, $\max(A_m, B_n) \xrightarrow{\mathbb{P}} \sigma_0$, and (4.4.7) follows.

To prove (4.4.8), it suffices to show that $q_{m,n}^{\dagger}(\sigma^*) - q_{m,n}^{\dagger}(\sigma_0) \xrightarrow{\mathbb{P}} 0$ since we have $q_{m,n}^{\dagger}(\sigma_0) - q_{m,n}^{*}(\sigma_0) \xrightarrow{\mathbb{P}} 0$ by applying Theorem 4.24.

For any $\sigma_1, \sigma_2 \in \mathbb{R}$, according to Theorem 4.5 and following the proof of Theorem 4.6, given $\boldsymbol{Z}_{m+n}$, it is not difficult to show that

$$\left|\mathcal{H}_{m,n}^{\dagger}(\sigma_1) - \mathcal{H}_{m,n}^{\dagger}(\sigma_2)\right| \leq 4\left(\left|Q_{\boldsymbol{Z}_{m+n}}(\varepsilon)\right| + \left|Q_{\boldsymbol{Z}_{m+n}}(1-\varepsilon)\right|\right)|\sigma_1 - \sigma_2| + 2\varepsilon,$$

so does their quantiles, i.e.

$$\left|q_{m,n}^{\dagger}(\sigma_1) - q_{m,n}^{\dagger}(\sigma_2)\right| \leq 4\left(\left|Q_{\boldsymbol{Z}_{m+n}}(\varepsilon)\right| + \left|Q_{\boldsymbol{Z}_{m+n}}(1-\varepsilon)\right|\right)|\sigma_1 - \sigma_2| + 2\varepsilon,$$

When the support of F and G are both bounded, $\sup_{m,n}\left(\left|Q_{\boldsymbol{Z}_{m+n}}(\varepsilon)\right| + \left|Q_{\boldsymbol{Z}_{m+n}}(1-\varepsilon)\right|\right)$ is bounded for any ε. When the support of E is unbounded, we can always find ε that is sufficiently small such that $e(E^{-1}(1-\varepsilon)) > 0$. By continuity of e, we can find $c_1, c_2 > 0$ such that $e(x) > c_2$ for any $x \in (E^{-1}(1-\varepsilon) - c_1, E^{-1}(1-\varepsilon) + c_1)$. Then by considering Dvoretzky–Kiefer–Wolfowitz inequality, we have

$$\begin{aligned}&\mathbb{P}\{Q_{\boldsymbol{Z}_{m+n}}(1-\varepsilon) > E^{-1}(1-\varepsilon) + c_1\} \\ &\qquad\qquad \leq \mathbb{P}\left(\sup_x |E_{m+n}(x) - E(x)| > c_1 c_2\right) \leq 2e^{-2(m+n)c_1^2 c_2^2}.\end{aligned}$$

Therefore for any $\delta > 0$, we can pick c_2 and N_0 such that $2e^{-2N_0 c_1^2 c_2^2} < \delta$. Therefore, for any $\delta > 0$, there exists $C > 0$ and N_0 such that

$$\sup_{m+n > N_0} \mathbb{P}\{Q_{\boldsymbol{Z}_{m+n}}(1-\varepsilon) > C\} < \delta, \tag{4.C.25}$$

which holds true regardless of the support set of F and G. Similarly, we can also

show that (4.C.25) holds true with respect to $|Q_{\boldsymbol{Z}_{m+n}}(\varepsilon)|$. Then for any $\eta > 0$,

$$\mathbb{P}\left(\left|q^{\dagger}_{m,n}(\sigma^*) - q^{\dagger}_{m,n}(\sigma_0)\right| > \eta\right) \leq \mathbb{P}(4C|\sigma^* - \sigma_0| + 2\varepsilon > \eta) + \mathbb{P}\left\{|Q_{\boldsymbol{Z}_{m+n}}(\varepsilon)| + |Q_{\boldsymbol{Z}_{m+n}}(1-\varepsilon)| > C\right\}.$$

Then we pick $\varepsilon < \eta$, $C = |E^{-1}(1-\varepsilon)| + |E^{-1}(\varepsilon)|$, then by (4.4.7) and (4.C.25), we can find N_0, such that when $m, n > \delta$, $\mathbb{P}\{|Q_{\boldsymbol{Z}_{m+n}}(\varepsilon)| + |Q_{\boldsymbol{Z}_{m+n}}(1-\varepsilon)| > C\} < \delta$ and $\mathbb{P}(4C|\sigma^* - \sigma_0| + 2\varepsilon > \eta) < \delta$. Therefore, we have

$$\mathbb{P}\left(\left|q^{\dagger}_{m,n}(\sigma^*) - q^{\dagger}_{m,n}(\sigma_0)\right| > \eta\right) \leq \delta,$$

for sufficiently large m and n, i.e. $q^{\dagger}_{m,n}(\sigma^*) - q^{\dagger}_{m,n}(\sigma) \xrightarrow{\mathbb{P}} 0$. Therefore, the result follows. □

In order to prove Theorem 4.36, we need to start with the following lemma.

Lemma 4.44 (Lemma 3 of Abrahamson 1967)**.** *Let $Z_1^{(i)}, Z_2^{(i)}, \ldots$ be independent sequences of independent random variables $Z^{(i)}, i = 1, 2$. Let the mean of the first l observations in the ith sample be $\bar{Z}_l^{(i)}$. Let $\varphi_i(t) = \mathbb{E}(e^{tZ^{(i)}})$.*

Suppose there exists a positive number τ in the interior of

$$T_\eta = \{t : \varphi_1(t/\eta)\varphi_2((-t)/(1-\eta)) < \infty\} \tag{4.C.26}$$

such that, for a given number $\epsilon > 0$, $\epsilon > \mathbb{E}[Z^{(1)} - Z^{(2)}]$,

$$\varphi_1'(\tau/\eta)/\varphi_1(\tau/\eta) - \varphi_2'(-\tau/(1-\eta))/\varphi_2(-\tau/(1-\eta)) = \epsilon, \tag{4.C.27}$$

and let

$$\rho = e^{-\tau\epsilon}\left[\varphi_1(\tau/\eta)\right]^{\eta}\left[\varphi_2((-\tau)/(1-\eta))\right]^{1-\eta}. \tag{4.C.28}$$

Then

$$\mathbb{P}\left\{\bar{Z}_m^{(1)} - \bar{Z}_n^{(2)} \geq \epsilon\right\} \leq \rho^{n+m}, \tag{4.C.29}$$

and

$$\lim_{m,n\to\infty} \frac{1}{m+n} \log \mathbb{P}\left\{\bar{Z}_m^{(1)} - \bar{Z}_n^{(2)} \geq \epsilon\right\} = \log \rho \tag{4.C.30}$$

where $m, n \to \infty$ in such a way that $m/(m+n) = r/k = \eta$.

Proof of Theorem 4.36. Similarly to proving Theorem 3.41, in order to prove Theorem 4.36, we only need to prove that

$$\begin{aligned} H(F_\sigma, G_\sigma) &= \lim_{m,n\to\infty} \mathcal{H}_{m,n}(\sigma) \quad a.s., \\ -f_{\sigma,\eta}(q) &= \lim_{m,n\to\infty} \frac{1}{m+n} \log \mathbb{P}(\mathcal{H}_{m,n} \geq q) \quad a.s. \end{aligned} \tag{4.C.31}$$

The proof of the first equation of (4.C.31) is similar to proving the first equation of (3.A.33), thus will be omitted. Let us prove the second equation of (4.C.31).

For a given x, $G_n(x)$ is the mean of m independent copies of the random variables, which we denote as $V^{(1)}$ that follow the Bernoulli distribution. Then

$$\varphi_x(t) = \mathbb{E}(e^{tV^{(1)}}) = E(x)e^t + 1 - E(x) \tag{4.C.32}$$

Similarly, for a fixed x and q, $F_m(x+q)$ is also the mean of n independent copies of the random variables $V^{(2)}$, with

$$\mathbb{E}(e^{tV^{(2)}}) = E(x+q)e^t + 1 - E(x+q) = \varphi_x(t+q). \tag{4.C.33}$$

Let τ_1, τ_2 be functions of x and q implicitly defined by (4.5.3). Then by applying (4.C.30) of Lemma 4.44, we have

$$\begin{aligned} &\lim_{m,n\to\infty} \frac{1}{m+n} \log \mathbb{P}\{F_m(x-q) - G_n(x) \geq q\} = \\ &\qquad \eta \log(\varphi_x(\frac{\tau_1(x,q)}{\eta}) - q) + (1-\eta)\log(\varphi_x(\frac{-\tau_1(x,q)}{1-\eta})) - \tau_1(x,q)q, \\ &\lim_{m,n\to\infty} \frac{1}{m+n} \log \mathbb{P}\{G_n(x) - F_m(x+q) \geq q\} = \\ &\qquad \eta \log(\varphi_x(\frac{-\tau_1(x,q)}{\eta}) + q) + (1-\eta)\log(\varphi_x(\frac{\tau_1(x,q)}{1-\eta})) - \tau_2(x,q)q, \end{aligned}$$

when $\frac{m}{m+n} \to \eta$. Since, it is not difficult to show that $\mathbb{P}(\mathcal{H}_{m,n}(\sigma) > q) \geq \mathbb{P}(F_m(x-$

$q) - G_n(x) > q)$ and $\mathbb{P}(\mathcal{H}_{m,n}(\sigma) > q) \geq \mathbb{P}(G_n(x) - F_m(x+q) > q)$

$$\begin{aligned}
&\liminf_{m,n\to\infty, \frac{m}{m+n}\to\eta} \frac{1}{m+n} \log \mathbb{P}\left(\mathcal{H}_{m,n} \geq q\right) \geq \\
&\quad \sup_{0<x<1} \max\Big\{ \eta \log(\varphi_x(\frac{\tau_1(x,q)}{\eta}) - q) + (1-\eta)\log(\varphi_x(\frac{-\tau_1(x,q)}{1-\eta})) - \tau_1(x,q)q, \\
&\qquad\qquad \eta \log(\varphi_x(\frac{-\tau_1(x,q)}{\eta}) + q) + (1-\eta)\log(\varphi_x(\frac{\tau_1(x,q)}{1-\eta})) - \tau_2(x,q)q \Big\} \\
&\qquad\qquad = -f_{\sigma,\eta}(q)
\end{aligned}$$

Then it remains to show that $\liminf_{m,n\to\infty, \frac{m}{m+n}\to\eta} \frac{1}{m+n} \log \mathbb{P}\left(\mathcal{H}_{m,n} \geq q\right) \leq -f_{\sigma,\eta}(q)$. The remaining technical part is similar to proving Theorem 3.41, thus will be omitted. □

4.C.2 Results in Section 4.A

In order to prove Lemma 4.37, let us first introduce an auxiliary lemma.

Lemma 4.45. *When there are no ties in the sample, for two vertex projections $\lambda_{B_{j_0},t_0}$ and $\lambda_{B_{j_1},t_1}$, where $t_0, t_1 \in \{1,2\}$, if they are discordant, their corresponding omnidirectional jumps must also be discordant.*

Proof. To prove Lemma 4.45, it suffices to show that two concordant omnidirectional jumps β_{j_0} and β_{j_1} must have concordant vertex projections $\lambda_{B_{j_0},t_0}$ and $\lambda_{B_{j_1},t_1}$, $t_0, t_1 \in \{1,2\}$. For simplicity, let us assume that $\beta_{j_0} \preceq \beta_{j_1}$.

- Case 1. $t_0 = t_1$. This is trivial since if $\beta_{j_0} \preceq \beta_{j_1}$, $G_n(\beta_{j_0}) < G_n(\beta_{j_1})$.
- Case 2. $t_0 \neq t_1$. Since $G_n(\beta_{j_0}) < G_n(\beta_{j_1})$, it is obvious that $\lambda_{B_{j_0},2} \preceq \lambda_{B_{j_0},1}$. On the other hand, $G_n(\beta_{j_1}) \geq G_n(\beta_{j_1}) + \frac{1}{n}$, thus $G_n(w^{(1)}_{\beta_{j_1}} - \varepsilon_0, w^{(2)}_{\beta_{j_1}}) = G_n(w^{(1)}_{\beta_{j_1}}, w^{(2)}_{\beta_{j_1}} - \varepsilon_0) \geq G_n(\beta_{j_1}) - \frac{1}{n} \geq G_n(\beta_{j_1})$. Therefore we have $\lambda_{B_{j_0},1} \preceq \lambda_{B_{j_0},2}$.

□

Proof of Lemma 4.37. For any two omnidirectional jumps β_i and β_j of G_n, $G_n(\beta^{max}) = G_n(\beta_i) + G_n(\beta_j) - G_n(\beta^{min}) + \frac{1}{n}\sum_{i=1}^n \mathbb{1}(\beta^{min} \prec X_i \preceq \beta^{max})$, where $\beta^{min} = (\min(\beta_i^{(1)}, \beta_j^{(2)}), \min(\beta_i^{(2)}, \beta_j^{(2)}))$. When they are discordant, for $\varepsilon_0 \in \mathbb{R}$ defined in (4.A.1) and e_i, $i = 1,2$,

$$\begin{aligned}
&G_n(\beta^{max}) - G_n(\beta^{max} - \varepsilon_0 e_i) \\
&\quad \geq G_n(\beta_i) - G_n(\beta_i - \varepsilon_0 e_i) + G_n(\beta_j) - G_n(\beta_j - \varepsilon_0 e_i) - (G_n(\beta^{min}) - G_n(\beta^{min} - \varepsilon_0 e_i)).
\end{aligned}$$

Due to the continuity of F and G, $G_n(\beta^{min}) - G_n(\beta^{min} - \varepsilon_0 e_i) = 0$. Thus, $G_n(\beta^{max}) - G_n(\beta^{max} - \varepsilon_0 e_i) > 0$ for $i = 1, 2$. Therefore, β^{max} is also an omnidirectional jump of G_n. For two discordant projections $\lambda_{B_{j_0,t_0}}$ and $\lambda_{B_{j_1,t_1}}$, according to Lemma 4.45, their corresponding jumps β_{j_0} and β_{j_1} also must be discordant. Then we consider one of two cases

- Case 1. $G_n(\beta^{max}) - G_n(\beta_{j_0}) = G_n(\beta^{max}) - G_n(\beta_{j_1}) = \frac{1}{n}$ for $i = 1, 2$. Then there must exists an omnidirectional jump $\beta_{j_2} = (\max(\beta_{j_0}^{(1)}, \beta_{j_1}^{(2)}), \max(\beta_{j_0}^{(2)}, \beta_{j_1}^{(2)}))$ of G_n with corresponding projections

$$\lambda_{B_{j_2},2} = (\max(\lambda_{B_{j_0,1}}^{(1)}, \lambda_{B_{j_1,1}}^{(1)}), \max(\lambda_{B_{j_0,1}}^{(2)}, \lambda_{B_{j_1,1}}^{(2)})),$$
$$\lambda_{B_{j_2},4} = (\max(\lambda_{B_{j_0,2}}^{(1)}, \lambda_{B_{j_1,2}}^{(1)}), \max(\lambda_{B_{j_0,2}}^{(2)}, \lambda_{B_{j_1,2}}^{(2)})).$$

  Clearly $\lambda_{B_{j_2},4} \preceq (\max(\lambda_{B_{j_0,t_0}}^{(1)}, \lambda_{B_{j_1,t_1}}^{(1)}), \max(\lambda_{B_{j_0,t_0}}^{(2)}, \lambda_{B_{j_1,t_1}}^{(2)}))$, for $t_0, t_1 \in \{1, 2\}$.

- Case 2. $G_n(\beta^{max}) - G_n(\beta_{j_0}) \neq \frac{1}{n}$ or $G_n(\beta^{max}) - G_n(\beta_{j_1}) \neq \frac{1}{n}$. Without loss of generality, let us assume that $G_n(\beta^{max}) - G_n(\beta_{j_1}) > \frac{1}{n}$ and $\beta_{j_0}^{(1)} < \beta_{j_1}^{(1)}$. The case of $\beta_{j_0}^{(1)} > \beta_{j_1}^{(1)}$ will be omitted, as it is similar to that when $\beta_{j_0}^{(1)} < \beta_{j_1}^{(1)}$.

  (a) $G_n(\beta_{j_1}) < G_n(\beta_{j_0})$. Since no tie is appeared in the sample, there always exists $Y_{j_1'} \in \{x \in \mathbb{R}^2 : x \prec \beta^{max}, x \not\prec \beta_{j_1}\}$. Clearly, $Y_{j_1'}$ is also an omnidirectional jump which is discordant to β_{j_1}. Denote by $\tilde{\beta}_{j_1'} = (\beta_{j_0}^{(1)}, Y_{j_1'}^{(2)})$. $G_n(\beta^{max}) - G_n(\beta_{j_1}) > G_n(\tilde{\beta}_{j_1'}) - G_n(\beta_{j_1})$. Therefore, there must exists

$$Y_{j^*} \in \{x \in \mathbb{R}^2 : x \prec \beta^{max}, x \not\prec \beta_{j_1}\} \text{ such that } G_n(\tilde{\beta}_{j^*}) - G_n(\beta_{j_1}) = \frac{1}{n},$$

  where $\tilde{\beta}_{j^*} = (\beta_{j_0}^{(1)}, Y_{j^*}^{(2)})$. Since Y_{j^*} is discordant with β_{j_1}, $\tilde{\beta}_{j^*}$ must be an omnidirectional jump of G_n with $G_n(\tilde{\beta}_{j^*}) = G_n(\beta_{j_1}) + \frac{1}{n}$. Thus, we select $\beta_{j_2} = \tilde{\beta}_{j^*}$. Clearly, $\lambda_{B_{j_2},2} = (\beta_{j_1}^{(1)} + G_n(\beta_{j_1}), Y_{j^*}^{(2)} + G_n(\beta_{j_1})) \succeq \lambda_{B_{j_1},1}$ and $\lambda_{B_{j_2},4} \succeq \lambda_{B_{j_1},2}$. In addition, we have

$$Y_{j^*}^{(2)} + G_n(\beta_{j_1}) \leq \beta_{j_0}^{(2)} + G_n(\beta_{j_0}) - \frac{1}{n},$$

  due to that $Y_{j^*}^{(2)} < \beta_{j_0}^{(2)}$ and that $G_n(\beta_{j_1}) < G_n(\beta_{j_0})$. Therefore when $t_0 = 1, 2$, $t_1 = 1$, i.e. $\lambda_{B_{j_0},t_0}$ and $\lambda_{B_{j_0},1}$ are discordant, there exists $\lambda_{B_{j_2},2} \preceq$

$(\max(\lambda^{(1)}_{B_{j_0,t_0}},\lambda^{(1)}_{B_{j_1,t_1}}),\max(\lambda^{(2)}_{B_{j_0,t_0}},\lambda^{(2)}_{B_{j_1,t_1}}))$. When $t_0=1,2,\ t_1=2$, and there exists $\lambda_{B_{j_2},4}\preceq(\max(\lambda^{(1)}_{B_{j_0,t_0}},\lambda^{(1)}_{B_{j_1,t_1}}),\max(\lambda^{(2)}_{B_{j_0,t_0}},\lambda^{(2)}_{B_{j_1,t_1}}))$.

(b) $G_n(\beta_{j_1})>G_n(\beta_{j_0})$. Therefore $G_n(\beta^{max})-G_n(\beta_{j_0})>\frac{1}{n}$. This is similar to the case when $G_n(\beta_{j_1})<G_n(\beta_{j_0})$. Thus we can either select $\lambda_{B_{j_2},2}\succeq\lambda_{B_{j_0},1}$ or $\lambda_{B_{j_2},4}\succeq\lambda_{B_{j_0},2}$.

(c) $G_n(\beta_{j_1})=G_n(\beta_{j_0})$. The existence of j_2 also directly follows if $t_0=t_1$. If they are not equal, let us assume that $t_0=1$ and $t_1=2$, i.e. $\lambda_{B_{j_0},1}$ and $\lambda_{B_{j_1},2}$ are discordant. Following a similar argument, we can find β_{j_2} such that $\beta^{(1)}_{j_2}=\beta^{(1)}_{j_1}$ and $\beta^{(2)}_{j_0}>\beta^{(2)}_{j_2}>\beta^{(2)}_{j_1}$ with $G_n(\beta_{j_2})=G_n(\beta_{j_1})+\frac{1}{n}$. Then we have $\lambda_{B_{j_2},4}=(\beta^{(1)}_{j_2}+G_n(\beta_{j_1})-\frac{1}{n},\beta^{(2)}_{j_2}+G_n(\beta_{j_1})-\frac{1}{n})\succeq\lambda_{B_{j_2},2}$. Note that $\lambda_{B_{j_0},1}=(\beta^{(1)}_{j_0}+G_n(\beta_{j_0}),\beta^{(2)}_{j_0}+G_n(\beta_{j_0}))$, $\beta^{(2)}_{j_2}+G_n(\beta_{j_1})-\frac{1}{n}\leq\beta^{(2)}_{j_0}+G_n(\beta_{j_0})$. Thus we are able to find $\lambda_{B_{j_2},4}$ meet the requirement.

Therefore, the existence of $\lambda_{B_{j_2},t_2}$ follows. □

To prove Lemma 4.39, we first introduce the following auxiliary lemma.

Lemma 4.46. *For any $(x,y)\in\mathbb{R}^2$, if $\mathcal{V}(x,y)$ is not empty, there always exists $\lambda\in\tilde{\mathcal{V}}(x,y)$ and $\mathcal{I}_{j,t}$, such that $\lambda,(x,y)\in\mathcal{I}_{j,t}$.*

Proof. If $\mathcal{V}(x,y)$ is not empty, it is clear that $(x,y)\in\mathcal{I}_{j,t}$ must hold for some j and t. Since $(x,y)\in\mathcal{I}_{j,t}$, there must exists $(x_1,y_1),(x_2,y_2)\in\mathcal{I}_{j,t}$ such that

$$(x_1,y_1)\preceq(x,y)\preceq(x_2,y_2). \tag{4.C.34}$$

Therefore, there should at least exists λ, one vertices of $\mathcal{I}_{j,t}$, such that $\lambda\in\mathcal{V}(x,y)$. Additionally, since the set in (4.A.2) are non-overlapping, at least one vertices of $\mathcal{I}_{j,t}$ should further fall in $\tilde{\mathcal{V}}(x,y)$. Otherwise, for any $(x_0,y_0)\in\mathcal{I}_{j,t}$, $(x,y)\npreceq(x_2,y_2)$, which contradicts (4.C.34). □

Proof of Lemma 4.39. We consider two cases of $\lambda^*_B(A_l)$.

- Case 1. $\lambda^*_B(A_l)$ is determined by (4.A.3). Therefore, $\lambda_{B_{j_0,t_0}}\in\tilde{\mathcal{V}}(\lambda_{A_l})$, for some j_0 and $t_0=1$ or 2. By the definition of $\tilde{\mathcal{V}}(\lambda_{A_l})$, we have $\lambda^*_B(A_l)\preceq\lambda_{A_l}$.

  (a) $t_0=1$. It is trivial to show that region $\mathcal{I}_{j_0,1}$ contains λ_{A_l}, which corresponds to the surface with the normal direction parallel to the axis

Oz. By the definition of E_l, the projection of E_l and A_l along the direction OE_0 should coincide. Therefore, the surface with projection corresponding to the region $\mathcal{I}_{j_0,3}$ should contain the point E_l. Therefore, we have $w^{(3)}_{E_l} = w^{(3)}_{B_{j_0,1}}$, where E_l is the intersection of line $\mathcal{L}_l$ and the planar curve G^c_n, defined in Lemma 4.15. Additionally, by the property of E_l, we have $|w^{(1)}_{E_l} - w^{(1)}_{A_l}| = |w^{(2)}_{E_l} - w^{(2)}_{A_l}| = |w^{(3)}_{E_l} - w^{(3)}_{A_l}|$, thus we have $\inf_{B\in G^c_n} \rho_\infty(A_l, B) = \rho_\infty(A_l, E_l) = |w^{(3)}_{A_l} - w^{(3)}_{B_{j_0,t_0}}|$.

(b) $t_0 = 2$. This is the case when either $\mathcal{I}_{j_0,1}$ or $\mathcal{I}_{j_0,2}$ contains λ_{A_l}. This can be further specified by checking whether $w^{(1)}_{B_{j_0,2}} - w^{(1)}_{A_l} \le w^{(2)}_{B_{j_0,2}} - w^{(2)}_{A_l}$ holds. If $w^{(1)}_{B_{j_0,2}} - w^{(1)}_{A_l} \le w^{(2)}_{B_{j_0,2}} - w^{(2)}_{A_l}$, then $\lambda_{B_{j_0,2}} \in \mathcal{I}_{j_0,1}$. Similarly, we have $w^{(1)}_{E_l} = w^{(1)}_{B_{j_0,2}}$. Thus $\inf_{B\in G^c_n} \rho_\infty(A_l, B) = \rho_\infty(A_l, E_l) = |w^{(1)}_{A_l} - w^{(1)}_{B_{j_0,t_0}}|$. The case when $\lambda_{B_{j_0,2}} \in \mathcal{I}_{j_0,2}$ is similar.

- Case 2. $\lambda^*_B(A_l)$ is determined by (4.A.5). Therefore, $\tilde{\mathcal{V}}(\lambda_{A_l})$ is not empty and $\forall\, j = 1,\dots,\nu$ and $t = 1,2$, $\lambda_{B_{j,t}} \notin \tilde{\mathcal{V}}(\lambda_{A_l})$. Our first step is to show that $\lambda^*_B(A_l)$ defined in (4.A.4) fall in the same region as λ_{A_l}, i.e. $\lambda^*_B(A_l), \lambda_{A_l} \in \mathcal{I}_{j,t}$ for some j and t. By Lemma 4.46, we could always find $\lambda \in \tilde{\mathcal{V}}(\lambda_{A_l})$, such that $\lambda, \lambda_{A_l} \in \mathcal{I}_{j,t}$. Thus, it suffices to show that if $\lambda_{j_1,4} \neq \lambda_{j_2,4} \in \tilde{\mathcal{V}}(\lambda_{A_l})$, $\rho_\infty(\lambda_{j_1,4}, \lambda_{A_l}) > \rho_\infty(\lambda_{j_2,4}, \lambda_{A_l})$, then $\lambda_{j_1,4} \notin \mathcal{I}_{j,t}$. This can be shown graphically by Figure 4.13. Since $\rho_\infty(\lambda_{j_1,4}, \lambda_{A_l}) > \rho_\infty(\lambda_{j_2,4}, \lambda_{A_l})$, it is clear that λ falls

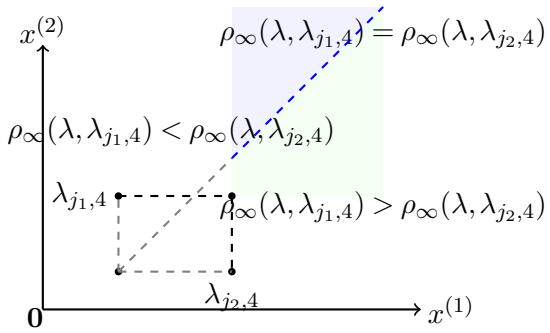


Figure 4.13: Regions where $\rho_\infty(\lambda, \lambda_{j_1,4}) > \rho_\infty(\lambda, \lambda_{j_2,4})$ or $\rho_\infty(\lambda, \lambda_{j_1,4}) < \rho_\infty(\lambda, \lambda_{j_2,4})$ for $\lambda \succeq \lambda_{j_1,4}, \lambda_{j_2,4}$, with the boundary $\rho_\infty(\lambda, \lambda_{j_1,4}) = \rho_\infty(\lambda, \lambda_{j_2,4})$ shown as a blue dashed line.

in the light green region of Figure 4.13. Additionally, the equidistant line $\rho_\infty(\lambda, \lambda_{j_1,4}) = \rho_\infty(\lambda, \lambda_{j_2,4})$ is of 45 degrees. By the shape of $\mathcal{I}_{\cdot,2}$ and $\mathcal{I}_{\cdot,3}$, it is clear that $\lambda_{j_1,4} \notin \mathcal{I}_{j,t}$. Thus $\lambda^*_B(A_l)$ defined in (4.A.4) should be in the same region as λ_{A_l}. By the shape of regions in (4.A.2), it is also easy to show that

$\lambda_B^*(A_l)$ in (4.A.5) also stays in the same set in (4.A.2) as λ_{A_l}. Therefore, (4.A.6) can also prove to be true by using the same step as in Case 1(b)).

In summary, (4.A.6) holds. □

4.D Accuracy and Speed Comparisons

In this appendix, we compare the speed and accuracy of the C++ implementation of the projection and the transformation method in Section 4.3.1 for computing the Hausdorff distance, $H(F_m^c, G_n^c)$. The reported CPU times are obtained running the related C++ code on a machine with a 2.20 GHz Intel Core i7-14650HX processor with 16GB RAM, running Windows 11.

We separately consider cases $n = m =$20, 50, 100, 500, 1000, 5000, 10000, and 50000. In each case of sample sizes, we generate two random samples from the uniform distribution $U(0,1)$ and then compute the $H(F_m^c, G_n^c)$. To further evaluate the accuracy, we further round each observation in both samples to 6 decimal places. Thus, although the statistical value $H(F_m^c, G_n^c)$ is sample dependent (therefore random), in the view of Lemma 4.23, the real value of $H(F_m^c, G_n^c)$ must be an integer multiple of the value 10^{-6}. With this available, we provide the run time for computing $H(F_m^c, G_n^c)$ 1000 times, as well as their precision, in Table 4.2. Clearly,

$m = n$	Algorithm	$H(F_m^c, G_n^c)$	Rel.Err	CPU time	Relative time
20	P	0.15000000000000000	0	0.0057	0.98
	Tr	0.15000000000000000	0	0.0058	1
50	P	0.08000000000000000	0	0.0098	1.04
	Tr	0.07999999999999990	1.22×10^{-15}	0.0094	1
100	P	0.04000000000000000	0	0.0184	1.18
	Tr	0.04000000000000000	0	0.0157	1
500	P	0.02526300000000000	0	0.2455	2.47
	Tr	0.02526300000000090	0	0.0994	1
1,000	P	0.01600000000000000	0	0.9185	3.91
	Tr	0.01600000000000020	1.27×10^{-14}	0.2348	1
5,000	P	0.00901700000000005	5.33×10^{-15}	20.0851	13.9
	Tr	0.00901699999999918	9.10×10^{-14}	1.4468	1
10,000	P	0.00639999999999996	6.22×10^{-15}	83.3945	26.0
	Tr	0.00640000000000377	5.89×10^{-13}	3.2110	1
50,000	P	0.00165999999999999	6.00×10^{-15}	2002.7778	87.9
	Tr	0.00165999999997949	1.24×10^{-11}	22.7778	1

Table 4.2: The computed distance $H(F_m^c, G_n^c)$ with the corresponding relative error and the CPU time for 1000 times repeated evaluations using C++ implementation of projection (P) and Transformation (Tr) approach

the projection approach usually provide 16 correct digits, in contrast to the transformation approach, which typically provides 14 correct digits. This is at a cost of higher run time, since the former approach is 1-87 times slower than the latter, which further grows as sample sizes increase.

Chapter 5

Conclusions and Further Developments

5.1 Summary of Potential Extensions

In Chapter 2, we have given recursive formulae (2.2.11) and (2.2.23) for the computing of the exact p-values of the KS tests and Kuiper tests with arbitrary weights for data samples coming from discontinuous distributions. In principle, these recursive formulae potentially can be generalized to the multivariate KS or Kuiper tests with appropriate definitions.

We have further shown that the weight function suggested by Büning (2001) can significantly improve the power of the KS test when samples come from purely discrete distributions. The use of some particular weight functions of the KS test has been thoroughly explored for continuous data samples; however, it remains to be seen how to systematically select weight functions in the more general case, which will be interesting to investigate in the future.

In addition, we have shown that the recursive formula (2.2.23) is efficient only if the sample sizes m and n have a small least common multiple and is extremely time-consuming if m and n are coprime. This computational inefficiency can be addressed by two possible directions: one is to give a bound of the approximation error arising from replacing $m' \approx m, n' \approx n$ so that it can be reasonably controlled, the other is to further give a universal recursion formula without relying on the least common multiple.

Potential extensions arising from Chapter 3 are given in Section 5.2.

In Chapter 4, we introduced the two-sample statistic based on the Hausdorff

distance in the general case of dimension $k \geq 1$ and developed a numerical method to compute it when $k \leq 2$. However, the computation complexity of the latter method increases with k increasing, since it relies on searching and sorting the onmidirectional jumps in the empirical cdfs F_m and G_n, whose numbers grow at rates m^k and n^k respectively. Therefore, a potential further extension would be to find an efficient method that evaluates the distance between empirical cdfs in high dimensions with lower computational complexity. Furthermore, it would be informative to conduct more systematic power comparisons against the existing test statistics in high dimensions.

Another interesting extension of the results in Chapter 4 would be to consider possibly dependent data and develop statistics for testing (conditional) independence based on the Hausdorff distance. In particular, Kim et al. (2022) propose a general framework for conditional independence testing based on local permutations in the sample. This line of work has been extended further by Neykov et al. (2024), who construct Wasserstein-type statistics for conditional independence and obtain several fruitful results, where smoothness assumptions play a crucial role. We believe that this framework could be further generalized to construct independence test statistics based on the Hausdorff metric, relying on the Lipschitz continuity properties established in Theorem 4.6.

5.2 New Metrics and Statistics

In Chapter 3, we have investigated the properties of the Hausdorff statistic and its scaled version $\mathcal{H}_n(\sigma)$. We have particularly highlighted its connections to the KS statistic and shown in Theorem 3.17 that it can be rewritten in the form of the KS test. We would like to highlight that this approach can be extended to introducing a new metric in analogy to the Kuiper test. Following (3.2.7), which is the functional form of the Hasudorff metric, we define the distance H' between cdfs F and F_n as

$$H'(F, F_n) = \sup_y \inf_x \max\left(|x-y|, F(x) - F_n(y)\right) + \sup_x \inf_y \max\left(|x-y|, F_n(y) - F(x)\right). \tag{5.2.1}$$

It is not difficult to show that $H'(F, F_n)$ can be expressed as

$$H'(F^c, F_n^c) = \sup_t [K_n(t) - K(t)] - \inf_t [K_n(t) - K(t)],$$

similarly as in Theorem 3.17, where K_n and K are defined in Example 3.16. Since Abrahamson (1967) has shown that the Bahadur exact slope of the Kuiper test is higher than that of the KS test, following the similar logic, one should expect that the statistic based on the metric H' may have higher Bahadur efficiency compared with the Hausdorff metric, i.e., it has higher power.

Based on the Proposition 3.34, the values of $\mathcal{H}_n(\sigma)$ can be interpreted as the vertical side of the largest rectangle among all rectangles with ratio of sides equal to σ, inscribed between F^c and F_n^c. Consequently, the values of $\mathcal{H}_n(\sigma)$ across different σ are not directly comparable. A more natural normalisation is to consider $\frac{1}{\sigma}\mathcal{H}_n^2(\sigma)$ or $\frac{1}{\sqrt{\sigma}}\mathcal{H}_n(\sigma)$ which both represent the area of that largest rectangle. In particular, $\frac{1}{\sqrt{\sigma}}\mathcal{H}_n(\sigma)$ can be meaningfully compared across different σ.

Since different choices of σ emphasise different aspects of the deviation between F^c and F_n^c, a statistic that aggregates all such information can be defined as

$$H_{max}^p = \max_{\sigma>0}\left\{w(\sigma)\left[\frac{1}{\sqrt{\sigma}}\mathcal{H}_n(\sigma)\right]^p\right\},\ H_{int}^p = \int_0^\infty w(\sigma)\left[\frac{1}{\sqrt{\sigma}}\mathcal{H}_n(\sigma)\right]^p d\sigma,$$

where $w(\sigma)$ is an appropriately chosen weight function on $\sigma \in (0,\infty)$ guaranteeing that both statistics are well defined, i.e., $H_{max}^p < \infty$ and $H_{int}^p < \infty$. More importantly, since $\frac{1}{\sqrt{\sigma}}\mathcal{H}_n(\sigma)$ represents the area, it is not difficult to show that any rescaling with respect to F and F_n does not affect the p-values of H_{max}^p and H_{int}^p. When $w(\sigma) \equiv 1$, the value H_{max}^2 has the simple geometric interpretation of being the area of the largest rectangle inscribed between F^c and F_n^c.

Furthermore, combining the reasoning based on Section 3.3.2 and the rule (3.3.11), when $\psi_1 \approx \psi_2 \approx \psi$, (3.3.11) reduces to $\sigma = f(F^{-1}(\psi))$. Substituting this reparametrisation into H_{int}^p yields

$$H_{int}^p = \int_0^1 \tilde{w}(\psi)\left[\frac{1}{\sqrt{f\left(F^{-1}(\psi)\right)}}\mathcal{H}_n\left(f\left(F^{-1}(\psi)\right)\right)\right]^p d\psi,$$

where $\tilde{w}(\psi)$ is a weight function with respect to $\psi \in (0,1)$.

Intuitively, $\frac{1}{\sqrt{f(F^{-1}(\psi))}}\mathcal{H}_n\left(f\left(F^{-1}(\psi)\right)\right)$ is the statistic that are sensitive around the ψ quantile of F. Therefore, by choosing $\tilde{w}(\psi)$ to take larger values when ψ is close to 0 or 1, one can construct a statistic that is simultaneously sensitive to deviations in both the left and right tails of the null distribution, regardless of the

shape of F. This stands in contrast to the original $\mathcal{H}_n(\sigma)$, for which it is impossible to tune σ so as to achieve simultaneous left- and right-tail sensitivity when F is purely concave or purely convex. However, all of the above statistics effectively introduce new metrics on the space of distribution functions and therefore require systematic investigation. For instance, one can show that when $w(\sigma) \equiv 1$, H^2_{max} is well defined (i.e. finite) for any distribution function F satisfying

$$\int_0^{+\infty} [1 - F(t)]dt + \int_{-\infty}^0 F(t)dt < \infty.$$

However, in general, H^2_{max} need not be finite for arbitrary F, and it may fail to define a well-defined metric on the full class of distributions.

In Chapter 4, we have also highlighted the connection between the Hausdorff metric and the Levy-Prokhorov metric. Another interesting investigation will be to extend the existing methods and theoretical results for the Hausdorff metric to the statistic constructed based on the Levy-Prokhorov metric, i.e. $\rho_{LP}(\mu_{\boldsymbol{X}_m}, \mu_{\boldsymbol{Y}_n})$, which again is scale-dependent but robust to mild perturbations. We should highlight that Alexander (1974) has investigated a statistic based on the Prokhorov metric and applying rank transformation. However, as we have similarly argued in Remark 3.24 and Section 4.4, such a rank-statistic is less appealing since it does not take into account the information contained in each observation and violates many robustness properties. We still believe the approach to tune scale rather than applying the ranking transformation gives potential to develop more powerful tests against the Wasserstein test and more flexibility to develop locally sensitive tests.

We also believe this statistic will be particularly powerful for (two-sample) high-dimensional statistics since it does not rely on any specific definitions of the empirical cdf and thus is independent of the relative ordering in $\mathbb{R}^k$. In addition, Garel and Massé (2009) have shown in their Lemma 5 that the exact Prokhorov distance between finite-support probability measures can be obtained by solving an optimization problem through the simplex method. The latter requires the knowledge of the distances between different data points and thus the complexity does not grow as the dimension grows, in contrast to the H statistic.

For the purpose of computing the p-values of $\rho_{LP}(\mu_{\boldsymbol{X}_m}, \mu_{\boldsymbol{Y}_n})$, one will need the permutation version $\rho_{LP}(\mu_{\boldsymbol{X}^\dagger_m}, \mu_{\boldsymbol{Y}^\dagger_n})$ and the theory regarding their p-value differ-

ences, similar to Theorem 4.24. It may be possible to extend Theorem 4.24 to the univariate $\rho_{LP}(\mu_{\boldsymbol{X}_m}, \mu_{\boldsymbol{Y}_n})$ through the following route. First, we consider the class $\mathfrak{C}_k$, which denotes the collection of unions of finite (at most k) intervals, then one can extend the weak convergence in (4.C.13) of Lemma 4.42 over the indicator functions $\mathbb{1}_C$, $C \in \mathfrak{C}_k$, since $\mathfrak{C}_k$ has finite Vapnik–Chervonenkis dimension. Therefore one can extend Theorem 4.24 to the statistic $\rho_{LP}(\mu_{\boldsymbol{X}_m}, \mu_{\boldsymbol{Y}_n}; \mathfrak{C}_k)$ and its permutation version. Second, if it is possible to establish a bound

$$\sqrt{\frac{mn}{m+n}}\,|\rho_{LP}(\mu_{\boldsymbol{X}_m}, \mu_{\boldsymbol{Y}_n}) - \rho_{LP}(\mu_{\boldsymbol{X}_m}, \mu_{\boldsymbol{Y}_n}; \mathfrak{C}_k)| < \alpha(k),\ \forall\, m, n$$

where $\alpha(k) \to 0$ as $k \to \infty$ under certain regularity conditions of the underlying distributions F and G, Theorem 4.24 can then be extended to $\rho_{LP}(\mu_{\boldsymbol{X}_m}, \mu_{\boldsymbol{Y}_n})$.

Furthermore, we emphasise that, unlike the Hausdorff metric, which is only quasi-convex with respect to the probability measures, the Prokhorov metric is genuinely convex. This makes Prokhorov-based constructions particularly attractive in global sensitivity analysis and variable-importance measures, where convexity underlies several other desirable properties Borgonovo et al. (2025a). In addition, as already mentioned, the Prokhorov distance could be tuned to emphasise different features of the distributions, in close analogy with the Hausdorff metric. Therefore, we expect that, in the context of variable-importance analysis, Prokhorov-based indices would systematically reveal how the distribution of the output responds to perturbations in each input.

Bibliography

Abrahamson, I.G., 1967. Exact Bahadur Efficiencies for the Kolmogorov-Smirnov and Kuiper One- and Two-Sample Statistics. The Annals of Mathematical Statistics 38, 1475 – 1490.

Albrecher, H., Beirlant, J., Teugels, J.L., 2017. Reinsurance: Actuarial and Statistical Aspects. John Wiley & Sons.

Alexander, C.H., 1974. Statistical Tests Based on the Levy and Prohorov Metrics. Ph.D. thesis. North Carolina State University. Raleigh, NC.

Algolytics Technologies, 2022. **AdvancedMiner**, Version 3.3.11. Warszawa, Poland.

Alzubaidi, A., Kalita, J., 2016. Authentication of Smartphone Users Using Behavioral Biometrics. IEEE Communications Surveys & Tutorials 18, 1998–2026.

Anderson, T.W., 1960. A Modification of the Sequential Probability Ratio Test to Reduce the Sample Size. The Annals of Mathematical Statistics , 165–197.

Anderson, T.W., Darling, D.A., 1952. Asymptotic Theory of Certain "Goodness of fit" Criteria Based on Stochastic Processes. The Annals of mathematical statistics , 193–212.

Apache Software Foundation, 2024. **Commons Math**: The Apache Commons Mathematics Library, Version 4.0-SNAPSHOT. Wakefield, Massachusetts.

Archibald, A., 2015. **Kuiper**: Kuiper test and other tools from circular statistics. URL: `http://github.com/aarchiba/kuiper`.

Arnold, T.B., Emerson, J.W., 2022. **dgof**: Discrete Goodness-of-Fit Tests. URL: `https://CRAN.R-project.org/package=dgof`. R package version 1.4.

Arsioli, B., Dedin, P., 2020. Machine Learning Applied to Multifrequency Data in Astrophysics: Blazar Classification. Monthly Notices of the Royal Astronomical Society 498, 1750–1764.

Avella-Medina, M., 2021. Privacy-preserving Parametric Inference: A Case for Robust Statistics. Journal of the American Statistical Association 116, 969–983.

Bahadur, R.R., 1971. Some Limit Theorems in Statistics. SIAM.

del Barrio, E., Cuesta-Albertos, J.A., Matrán, C., Csörgö, S., Cuadras, C.M., de Wet, T., Giné, E., Lockhart, R., Munk, A., Stute, W., 2000. Contributions of Empirical and Quantile Processes to the Asymptotic Theory of Goodness-of-fit Tests. Test 9, 1–96.

del Barrio, E., Cuesta-Albertos, J.A., Matran, C., Rodriguez-Rodriguez, J.M., 1999. Tests of Goodness of Fit Based on the L_2-Wasserstein Distance. The Annals of Statistics 27, 1230–1239.

Beer, G., 1984. The Approximation of Real Functions in the Hausdorff Metric. Houston J. Math. 10, 325–338.

Bezanson, J., Edelman, A., Karpinski, S., Shah, V.B., 2024. Julia: A Fresh Approach to Numerical Computing, Version 1.10. URL: `https://julialang.org`.

de Bezieux, H.R., 2024. **Ecume**: Equality of 2 (or k) Continuous Univariate and Multivariate Distributions. URL: `https://CRAN.R-project.org/package=Ecume`. R package version 0.9.2.

Biswas, M., Mukhopadhyay, M., Ghosh, A.K., 2014. A Distribution-free Two-sample Run Test Applicable to High-dimensional Data. Biometrika 101, 913–926.

Bloch, I., Atif, J., 2016. Defining and Computing Hausdorff Distances Between Distributions on the Real Line and on the Circle: Link Between Optimal Transport and Morphological Dilations. Mathematical Morphology-Theory and Applications 1.

Borgonovo, E., Figalli, A., Ghosal, P., Plischke, E., Savaré, G., 2025a. Convexity and measures of statistical association. Journal of the Royal Statistical Society Series B: Statistical Methodology 87, 1281–1304.

Borgonovo, E., Figalli, A., Plischke, E., Savaré, G., 2025b. Global Sensitivity Analysis via Optimal Transport. Management Science 71, 3809–3828.

Borrajo, M., González-Manteiga, W., Martínez-Miranda, M., 2024. Goodness-of-fit test for Point Processes First-order Intensity. Computational Statistics & Data Analysis 194, 107929.

Boulesteix, A.L., Strobl, C., 2007. Maximally Selected Chi-squared Statistics and Non-monotonic Associations: An Exact Approach Based on Two Cutpoints. Computational Statistics & Data Analysis 51, 6295–6306.

Brown, M., 2011. **KSNstat**: Generalised Smirnov Two-sample Homogeneity Tests. URL: `https://www.maplesoft.com/applications/view.aspx?SID=121124`. MAPLE package Software version MAPLE 15.

Büning, H., 2001. Kolmogorov-Smirnov- and Cramér–von Mises Type Two-sample Tests With Various Weight Functions. Communications in Statistics - Simulation and Computation 30, 847–865.

Canner, P.L., 1975. A Simulation Study of One- and Two-Sample Kolmogorov-Smirnov Statistics with a Particular Weight Function. Journal of the American Statistical Association 70, 209–211.

Chavent, M., 2004. A hausdorff distance between hyper-rectangles for clustering interval data, in: Banks, D., McMorris, F.R., Arabie, P., Gaul, W. (Eds.), Classification, Clustering, and Data Mining Applications, Springer Berlin Heidelberg, Berlin, Heidelberg. pp. 333–339.

Chesneau, C., Dewan, I., Doosti, H., 2016. Nonparametric Estimation of a Quantile Density Function by Wavelet Methods. Computational Statistics & Data Analysis 94, 161–174.

Chhatbar, P.Y., Francis, J.T., 2013. Towards a Naturalistic Brain-Machine Interface: Hybrid Torque and Position Control Allows Generalization to Novel Dynamics. PLOS ONE 8, 1–9.

Chung, E., Romano, J.P., 2013. Exact and Asymptotically Robust Permutation Tests. Annals of Statistics 41, 484–507.

Cox, N.J., 1998. **CIRCSTAT**: Stata modules to calculate circular statistics. URL: `https://ideas.repec.org/c/boc/bocode/s362501.html`. statistical Software Components S362501 Stata module.

Cramér, H., 1928. On the Composition of Elementary Errors: First Paper: Mathematical Deductions. Scandinavian Actuarial Journal 1928, 13–74.

Cuturi, M., 2013. Sinkhorn Distances: Lightspeed Computation of Optimal Transport. Advances in neural information processing systems 26.

Cytel Inc., 2019. StatXact, Version 12. Cambridge, MA.

Dimitrova, D.S., Kaishev, V.K., Tan, S., 2020. Computing the Kolmogorov-Smirnov Distribution When the Underlying CDF is Purely Discrete, Mixed, or Continuous. Journal of Statistical Software 95, 1–42.

Dowd, C., 2020. A New ECDF Two-Sample Test Statistic. arXiv preprint arXiv:2007.01360 .

Dowd, C., 2023. **twosamples**: Fast Permutation Based Two Sample Tests. URL: `https://CRAN.R-project.org/package=twosamples`. R package version 2.0.1.

Dudbridge, F., 2006. A Note on Permutation Tests in Multistage Association Scans. American Journal of Human Genetics 78, 1094–5; author reply 1096.

Eplett, W.J.R., 1982. The Distributions of Smirnov Type Two-Sample Rank Tests for Discontinuous Distribution Functions. Journal of the Royal Statistical Society: Series B (Methodological) 44, 361–369.

Faraway, J., Marsaglia, G., Marsaglia, J., Baddeley, A., 2021. **goftest**: Classical Goodness-of-Fit Tests for Univariate Distributions. URL: `https://CRAN.R-project.org/package=goftest`. R package version 1.2-3.

Feigelson, E., Babu, G.J., 2020. Beware the Kolmogorov-Smirnov test! Available at: `https://asaip.psu.edu/articles/beware-the-kolmogorov-smirnov-test/`. (Accessed: 01 Nov 2024).

Finner, H., Gontscharuk, V., 2018. Two-sample Kolmogorov–Smirnov-type Tests Revisited: Old and New tests in terms of Local Levels. The Annals of Statistics 46, 3014 – 3037.

Foutz, R.V., Birch, J.B., 1982. Tests for the Two-sample Problem Based on Empirical Probability Measures. Communications in Statistics - Theory and Methods 11, 1839–1853.

Friedman, J.H., Rafsky, L.C., 1979. Multivariate Generalizations of the Wald-Wolfowitz and Smirnov Two-sample Tests. The Annals of statistics , 697–717.

Garel, B., Massé, J.C., 2009. Calculation of the Prokhorov Distance by Optimal Quantization and Maximum Flow. AStA Advances in Statistical Analysis 93, 73–88.

Gretton, A., Sejdinovic, D., Strathmann, H., Balakrishnan, S., Pontil, M., Fukumizu, K., Sriperumbudur, B.K., 2012. Optimal Kernel Choice for Large-scale Two-sample Tests, in: Advances in Neural Information Processing Systems.

Hausdorff, F., 1914. Grundzüge der mengenlehre. volume 7. von Veit.

Henze, N., 1988. A Multivariate Two-sample Test Based on the Number of Nearest Neighbor Type Coincidences. The Annals of Statistics , 772–783.

Hilton, J.F., Mehta, C.R., Patel, N.R., 1994. An Algorithm for Conducting Exact Smirnov Tests. Computational Statistics & Data Analysis 17, 351–361.

Hirakawa, K., 1973. The Two-sample Kuiper test. TRU Mathematics 9, 99–118.

Hodges, J.L., 1958. The Significance Probability of the Smirnov Two-sample Test. Arkiv för Matematik 3, 469–486.

Hoeffding, W., 1952. The Large-Sample Power of Tests Based on Permutations of Observations. The Annals of Mathematical Statistics 23, 169 – 192.

Hundrieser, S., Klatt, M., Munk, A., Staudt, T., 2024. A Unifying Approach to Distributional Limits for Empirical Optimal Transport. Bernoulli 30, 2846–2877.

Huttenlocher, D., Klanderman, G., Rucklidge, W., 1993. Comparing Images Using the Hausdorff Distance. IEEE Transactions on Pattern Analysis and Machine Intelligence 15, 850–863.

IBM Corp, 2022. IBM SPSS Statistics for Windows, Version 29.0. IBM Corp., Armonk.

Jammalamadaka, S., Guerrier, S., Mangalam, V., 2020. A Two-sample Nonparametric Test for Circular Data - its Exact Distribution and Performance. Sankhya B 83.

Janssen, A., 2000. Global Power Functions of Goodness of fit Tests. The Annals of Statistics 28, 239–253.

Karimi, D., Salcudean, S.E., 2020. Reducing the Hausdorff Distance in Medical Image Segmentation With Convolutional Neural Networks. IEEE Transactions on Medical Imaging 39, 499–513.

Kim, I., Balakrishnan, S., Wasserman, L., 2020. Robust Multivariate Nonparametric Tests via Projection Averaging. The Annals of Statistics 48, 3417–3441.

Kim, I., Neykov, M., Balakrishnan, S., Wasserman, L., 2022. Local permutation tests for conditional independence. The Annals of Statistics 50, 3388–3414.

Kim, P., Jennrich, R.I., 1973. Tables of the Exact Sampling Distribution of the Two-sample Kolmogorov-Smirnov Criterion, D_{mn}, m $\leq$ n, in: Selected Tables in Mathematical Statistics, Providence, R. I.: American Mathematical Society. pp. 80–129.

Kim, P.J., 1969. On the Exact and Approximate Sampling Distribution of the Two Sample Kolmogorov-Smirnov Criterion D_{mn} , m $\leq$ n. Journal of the American Statistical Association 64, 1625–1637.

Kolmogorov, A., 1933. Sulla determinazione empirica di una lgge di distribuzione. Inst. Ital. Attuari, Giorn. 4, 83–91.

Komlós, J., Major, P., Tusnády, G., 1975. An Approximation of Partial Sums of Independent RV'-s, and the Sample DF. I. Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete 32, 111–131.

Kossobokov, V.G., Panza, G.F., 2020. A Myth of Preferred Days of Strong Earthquakes? Seismological Research Letters 91, 948–955.

Kuiper, N.H., 1960. Tests Concerning Random Points on a Circle, in: Proceedings of the Koinklijke Nederlandse Akademie van Wetenschappen. Series A, pp. 38–47.

Langrené, N., Warin, X., 2021. Fast multivariate empirical cumulative distribution function with connection to kernel density estimation. Computational Statistics & Data Analysis 162, 107267.

Lehmann, E.L., Romano, J.P., 2005. Testing Statistical Hypotheses. 3rd ed., Springer.

Lemeshko, B., Gorbunova, A., Lemeshko, S., Rogozhnikov, A., 2014. Solving Problems of Using Some Nonparametric Goodness-of-fit Tests. Optoelectronics, Instrumentation and Data Processing 50, 21–35.

Li, T., Pan, Q., Gao, L., Li, P., 2017. A Novel Simplification Method of Point Cloud with Directed Hausdorff Distance, in: 2017 IEEE 21st International Conference on Computer Supported Cooperative Work in Design (CSCWD), pp. 469–474.

Lin, D., White, J.M., Byrne, S., Bates, D., Noack, A., Pearson, J., Arslan, A., Squire, K., Anthoff, D., Papamarkou, T., Besançon, M., Drugowitsch, J., Schauer, M., other contributors, 2019. JuliaStats/Distributions.jl: A Julia Package for Probability Distributions and Associated Functions. URL: `https://doi.org/10.5281/zenodo.2647458`.

Lin, L.C.C., Hu, C.P., Takata, J., Li, K.L., Hui, C.Y., Kong, A.K.H., 2022. Investigation of the Timing and Spectral Properties of an Ultraluminous X-Ray Pulsar NGC 7793 P13. The Astrophysical Journal 924, 65.

Liu, X., Briol, F.X., 2024. On the robustness of kernel goodness-of-fit tests. arXiv preprint arXiv:2408.05854 .

Maag, U.R., Stephens, M.A., 1968. The V_{NM} Two-Sample Test. The Annals of Mathematical Statistics 39, 923 – 935.

Mason, D.M., Schuenemeyer, J.H., 1983. A Modified Kolmogorov-Smirnov Test Sensitive to Tail Alternatives. The Annals of Statistics 11, 933–946.

Matheron, G., 1974. Random Sets and Integral Geometry. Wiley Series in Probability and Mathematical Statistics, Wiley. URL: `https://books.google.co.uk/books?id=JVoFcgAACAAJ`.

McQuillan, A., Mazeh, T., Aigrain, S., 2013. Stellar Rotation Periods of the Kepler Objects of Interest: A Dearth of Close-in Planets Around Fast Rotators. The Astrophysical Journal Letters 775, L11.

Meintanis, S., Milošević, B., Jiménez–Gamero, M., 2024. Goodness–of–fit Tests Based on the Min–characteristic Function. Computational Statistics & Data Analysis 197, 107988.

Milbrodt, H., Strasser, H., 1990. On the Asymptotic Power of the Two-sided Kolmogorov-Smirnov Test. Journal of Statistical Planning and Inference 26, 1–23.

Moscovich, A., 2023. Fast Calculation of P-values for One-sided Kolmogorov-Smirnov Type Statistics. Computational Statistics & Data Analysis 185, 107769.

Mouli, S.C., Teixeira, L., Neville, J., Ribeiro, B., 2019. Deep Lifetime Clustering. arXiv preprint arXiv:1910.00547.

Mousavi, M., Davulcu, H., Ahmadi, M., Axelrod, R., Davis, R., Atran, S., 2022. Effective Messaging on Social Media: What Makes Online Content Go Viral?, in: Proceedings of the ACM Web Conference 2022, Association for Computing Machinery, New York, NY, USA. p. 2957–2966.

Neykov, M., Wasserman, L., Kim, I., Balakrishnan, S., 2024. Nearly minimax optimal wasserstein conditional independence testing. Information and Inference: A Journal of the IMA 13, iaae033.

Nikiforov, A.M., 1994. Algorithm AS 288: Exact Smirnov Two-Sample Tests for Arbitrary Distributions. Journal of the Royal Statistical Society. Series C (Applied Statistics) 43, 265–270.

Nikitin, Y., 1995. Asymptotic Efficiency of Nonparametric Tests. Cambridge University Press.

Novikov, A., Frishling, V., Kordzakhia, N., 1999. Approximations of Boundary Crossing Probabilities for a Brownian Motion. Journal of Applied Probability 36, 1019–1030.

Oracle Corporation, 2024. The Java Programming Language, Version SE 22.0.1. Austin, Texas.

Paltani, S., 2004. Searching for Periods in X-ray Observations Using Kuiper's Test-Application to the ROSAT PSPC Archive. Astronomy & Astrophysics 420, 789–797.

Pan, W., Tian, Y., Wang, X., Zhang, H., 2018. Ball Divergence: Nonparametric Two sample Test. Annals of statistics 46, 1109.

Peacock, J.A., 1983. Two-dimensional Goodness-of-fit Testing in Astronomy. Monthly Notices of the Royal Astronomical Society 202, 615–627.

Popov, A.T., 1999. Hausdorff Distance and Fractal Dimension Estimation by Mathematical Morphology Revisited., in: NSIP, pp. 90–94.

Pötzelberger, K., Wang, L., 2001. Boundary Crossing Probability for Brownian Motion. Journal of Applied Probability 38, 152–164.

Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P., 2007. Numerical Recipes: The Art of Scientific Computing. 3 ed., Cambridge University Press.

R Core Team, 2024. R: A Language and Environment for Statistical Computing, Version 4.4.0. R Foundation for Statistical Computing, Vienna, Austria.

Rachev, S., 1981. Hausdorff Metric Structure of the Space of Probability Measures. Journal of Soviet Mathematics 17, 2275–2288.

Rachev, S.T., 1984. Hausdorff Metric Construction in the Probability Measures Space. PLISKA Studia mathematica bulgarica 7, 152p–162p.

Reynkens, T., Verbelen, R., Beirlant, J., Antonio, K., 2017. Modelling Censored Losses Using Splicing: A Global Fit Strategy With Mixed Erlang and Extreme Value Distributions. Insurance: Mathematics and Economics 77, 65–77.

Rieder, H., 1982. Qualitative Robustness of Rank Tests. The Annals of Statistics 10, 205–211.

Rosenbaum, P.R., 2005. An Exact Distribution-free Test Comparing Two Multivariate Distributions based on Adjacency. Journal of the Royal Statistical Society Series B: Statistical Methodology 67, 515–530.

Ruan, Y., 2018. **kuiper.2samp**: Two-Sample Kuiper Test. URL: `https://CRAN.R-project.org/package=kuiper.2samp`. R package version 1.0.

Salman, I., Misirli, A.T., Juristo, N., 2015. Are Students Representatives of Professionals in Software Engineering Experiments?, in: 2015 IEEE/ACM 37th IEEE International Conference on Software Engineering, IEEE. pp. 666–676.

SAS Institute Inc., 2023. SAS/STAT Software, Version 9.4. Cary, NC: SAS Institute Inc.

Schilling, M.F., 1986. Multivariate Two-sample Tests Based on Nearest Neighbors. Journal of the American Statistical Association 81, 799–806.

Schneider, G., Chicken, E., Becvarik, R., 2023. **NSM3**: Functions and Datasets to Accompany Hollander, Wolfe, and Chicken - Nonparametric Statistical Methods, Third Edition. URL: `https://CRAN.R-project.org/package=NSM3`. R package version 1.18.

Schröer, G., Trenkler, D., 1995. Exact and Randomization Distributions of Kolmogorov-Smirnov Tests Two or Three Samples. Computational Statistics & Data Analysis 20, 185–202.

Sendov, B., Beer, G., 2012. Hausdorff Approximations. Mathematics and its Applications, Springer Netherlands.

Sidak, Z., Sen, P.K., Hajek, J., 1999. Theory of Rank Tests. Academic Press.

Smirnov, N., 1939. Sur les écarts de la courbe de distribution empirique. Recueil Mathématique (Matematiceskii Sbornik) N.S. 6, 3–26.

Smirnov, N.V., 1933. Estimate of Deviation Between Empirical Distribution Functions in Two Independent Samples. Bulletin Moscow University 2, 3–16.

Soni, P., Dewan, I., Jain, K., 2012. Nonparametric estimation of quantile density function. Computational Statistics & Data Analysis 56, 3876–3886.

StataCorp, 2023. Stata Statistical Software: Release 18. StataCorp LLC., College Station.

StatsDirect Ltd., 2024. StatsDirect, Version 4.0.3. Birkenhead, Wirral. URL: `http://www.statsdirect.com`.

Steck, G., 1971. Rectangle Probabilities for Uniform Order Statistics and the Probability that the Empirical Distribution Function lies Between Two Distribution Functions. The Annals of Mathematical Statistics 42, 1–11.

TIBCO Software Inc., 2010. S-PLUS Software, Version 8.2. Palo Alto, California.

van der Vaart, A.W., Wellner, J.A., 2023. Weak Convergence and Empirical Processes. Springer Series in Statistics. 2nd ed., Springer Cham.

Viehmann, T., 2021. Numerically More Stable Computation of the P-values for the Two-sample Kolmogorov-Smirnov Test. `arXiv:arXiv:2102.08037`.

Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S.J., Brett, M., Wilson, J., Millman, K.J., Mayorov, N., Nelson, A.R.J., Jones, E., Kern, R., Larson, E., Carey, C.J., Polat, I., Feng, Y., Moore, E.W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E.A., Harris, C.R., Archibald, A.M., Ribeiro, A.H., Pedregosa, F., van Mulbregt, P., SciPy 1.0 Contributors, 2020. **SciPy** 1.0: Fundamental Algorithms for Scientific Computing in Python. Nature Methods 17, 261–272.

Von Mises, R., 1931. Vorlesungen aus dem gebiete der angewandten mathematik: bd. Wahrscheinlichkeitsrechnung und ihre anvendung in der statistik und theoretischen physik. volume 1. F. Deuticke.

Waterloo Maple Inc., 2024. MAPLE, Version 2024.0. Maplesoft, A Division of Waterloo Maple Inc., Waterloo, Ontario. URL: `https://www.maplesoft.com`.

Wilcox, R.R., 2012. Comparing Two Groups, in: Introduction to Robust Estimation and Hypothesis Testing (Third Edition). Academic Press. chapter 5, pp. 137–213.

Wilcox, R.R., Schönbrodt, F.D., 2022. The **WRS** package for robust statistics in R. URL: `https://github.com/nicebread/WRS`. R package version 0.39.

Wolfram Research, Inc, 2024. Mathematica, Version 14.0. Champaign, Illinois.

Wyłomańska, A., Iskander, D.R., Burnecki, K., 2020. Omnibus Test for Normality Based on the Edgeworth Expansion. PLOS ONE 15, 1–36.

Yung, C., Çolak, G., Wang, W., 2008. Cycles in the IPO market. Journal of Financial Economics 89, 192–208.

Zaiontz, C., 2024. Real Statistics Resource Pack. URL: `www.real-statistics.com`. Excel package Release 9.1.

Zhang, W., Huang, J., Zhang, T., Tan, Q., 2022. A Risk-based Stochastic Model for Supporting Resources Allocation of Agricultural Water-energy-food System under Uncertainty. Journal of Hydrology 610, 127864.

Zhao, J., Qi, F., Ren, G., Xu, L., 2021. PhD Learning: Learning With Pompeiu-Hausdorff Distances for Video-Based Vehicle Re-Identification, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2225–2235.

Zolotarev, V.M., 1971. Estimates of the difference between distributions in the Lévy metric. Trudy Matematicheskogo Instituta Imeni VA Steklova 112, 224–231.

Zolotarev, V.M., 2011. Lévy Metric. Encyclopedia of Mathematics. URL: `http://encyclopediaofmath.org/index.php?title=L%C3%A9vy_metric&oldid=14153`. Accessed: 2023-10-13.