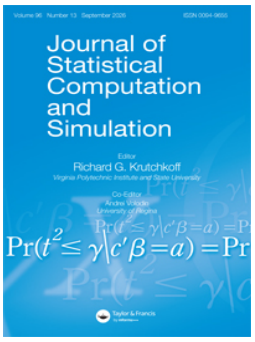

Journal of Statistical Computation and Simulation

ISSN: 0094-9655 (Print) 1563-5163 (Online) Journal homepage: www.tandfonline.com/journals/gscs20

Efficient exact calculation of *p*-values of the two-sample Kolmogorov-Smirnov and Kuiper tests

Dimitrina S. Dimitrova , Yun Jia & Vladimir K. Kaishev

JOURNAL OF STATISTICAL COMPUTATION AND SIMULATION
https://doi.org/10.1080/00949655.2026.2721410

Taylor & Francis by informa

Efficient exact calculation of *p*-values of the two-sample Kolmogorov-Smirnov and Kuiper tests

Dimitrina S. Dimitrova, Yun Jia and Vladimir K. Kaishev

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

ABSTRACT

We develop an efficient recurrence method for computing p-values of the two sample KS test. A similar method but for the unweighted KS test has been previously considered in the literature, without theoretical justification. We provide such justification., showing that the difference between the p-values of the permutation KS test and the KS test is asymptotically negligible; A flexible choice of the KS weight function is allowed, leading to better power; The method is extended to compute p-values of the two sample Kuiper test, thus covering the case of data with ties 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 provide an extensive overview of existing statistical software for computing KS and Kuiper p-values and compare with our proposed R functions `KS2sample` and `Kuiper2sample`, demonstrating their superior performance.

ARTICLE HISTORY
Received 18 February 2026
Accepted 7 August 2026

KEYWORDS
Kolmogorov-Smirnov test; Kuiper test; *p*-value; asymptotic distribution; statistical power

1. Introduction

In many areas of research and applied work, including economics, natural sciences, engineering, insurance, banking, and finance, large samples of data are collected and compared. A common problem is to test whether two samples come from the same, unspecified probability distribution. Among the many goodness-of-fit tests developed for this purpose (see e.g. [1,2]), two of the most widely used are the two-sample Kolmogorov-Smirnov (KS) test (see e.g. [3–6]) and the Kuiper test [7].

The two-sample KS statistic is the maximum absolute difference between the empirical distribution functions of the two samples. It is simple, intuitive, and widely used in applications, including astronomy [8], internet behaviour analysis [9], physics [10], and hydrology [11]. However, because the maximum difference is often attained near the centre of the distribution, the KS test is typically more sensitive there than in the tails.

The two-sample Kuiper statistic addresses this limitation by adding the largest positive and largest negative differences between the two empirical distribution functions. As a result, it is more equally sensitive across the support of the distribution, including both

CONTACT Vladimir K. Kaishev Vladimir.Kaishev.1@citystgeorges.ac.uk Bayes Business School, 106 Bunhill Row, London EC1Y 8TZ, UK

Supplemental data for this article can be accessed online at http://dx.doi.org/10.1080/00949655.2026.2721410.

the centre and the tails. It is also well suited to seasonal and circular data. The Kuiper test has therefore been used in a wide range of fields, including finance [12], genetics [13], neural networks [14], astronomy [15], brain-machine interfaces [16], and earthquake geology [17]. These applications highlight the importance of accurate and efficient methods for computing two-sample KS and Kuiper test statistics and their associated p-values.

This paper focuses on the computation of exact and asymptotic p-values for the two-sample KS and Kuiper tests when the sample sizes are finite and the observations may be continuous, discrete, or mixed. Accurate p-values are essential for testing the null hypothesis that two samples have a common underlying distribution. For continuous distributions, recurrence formulas for exact unweighted KS p-values were developed by Hodges [18], Kim [19], and Kim and Jennrich [20], while exact Kuiper formulas for equal sample sizes were given by Maag and Stephens [21] and Hirakawa [22]. These results rely on the distribution-free property of the tests under continuity. When ties are present, however, this property no longer holds in the same form, and the computation of exact p-values becomes more delicate. Related conditional formulas for the KS test were given independently by Nikiforov [3], Hilton et al. [23], and Schröer and Trenkler [24], but important theoretical and computational issues remain.

The present paper addresses these gaps. Its main contributions are as follows.

First, we provide a theoretical justification for computing KS p-values through a permutation formulation of the two-sample KS test. In particular, Theorem 2.1 shows that the difference between the critical values of the permutation KS test (c.f. (5)) and the usual empirical-distribution-function version of the KS test defined in (1) is asymptotically negligible for arbitrary underlying distributions. This result supports the recurrence method in formula (10) and its numerically enhanced version in formula (11).

Second, we give a detailed and self-contained description of the recurrence algorithm for exact KS p-values, including proofs and illustrative examples in the supplementary material. We generalize the method to allow an arbitrary weight function, which is important because the choice of weight can substantially affect the power of the KS test (see also [25]). We also implement two versions of the method: a faster version and a slower but more numerically stable version inspired by Viehmann [4]. Both are included in the R function `KS2sample`.

Third, we derive a closed-form expression for the asymptotic distribution of the two-sample KS statistic when ties are allowed; see formula (14) and Theorem 2.3. To the best of our knowledge, this asymptotic distribution has not previously been studied for tied observations. The resulting formula is both theoretically new and practically useful, providing an efficient alternative to exact computation for very large samples.

Fourth, we extend the recurrence approach to the two-sample Kuiper test. This yields exact Kuiper p-values for arbitrary continuous, discrete, or mixed observations, including tied data. Previous exact Kuiper formulas were limited to continuous observations and equal sample sizes [21,22]. We also develop a numerically enhanced version of the Kuiper algorithm and implement both versions in the R function `Kuiper2sample`.

Fifth, we compare the power of the Kuiper test with weighted versions of the KS test. In particular, we show that the Kuiper test can be a more powerful alternative for discrete observations, a setting that appears not to have been investigated previously.

Finally, we provide a detailed review and numerical comparison of existing software for computing two-sample KS and Kuiper p-values. Existing implementations often rely on approximate or resampling methods, focus mainly on the unweighted KS test, or do not adequately handle ties, especially for the Kuiper test. To address these limitations, we implement `KS2sample` and `Kuiper2sample` in the R package **KSgeneral**, thereby extending the package from one-sample to two-sample goodness-of-fit testing. We also provide Mathematica code for arbitrary-accuracy computation, which is used as a benchmark in our numerical comparisons.[1]

The paper is organized as follows. Section 2 introduces the weighted KS and Kuiper tests and describes the proposed methods for computing their p-values. Section 2.3 presents the asymptotic distribution of the two-sample KS statistic with ties. Section 3 gives two illustrative examples, including an application of `Kuiper2sample` to circular data with ties. Section 4 reviews existing software and compares it with the proposed implementations in terms of speed and accuracy. Section 5 concludes. Supplementary material contains proofs, algorithmic examples, power comparisons, and additional numerical results.

2. The two-sample KS and Kuiper test

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 $\boldsymbol{X}_m = (X_1, \dots, X_m)$ and $\boldsymbol{Y}_n = (Y_1, \dots, Y_n)$ 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:

$$\begin{aligned}
\Delta_{m,n} &= \sup_{x\in\mathbb{R}} |F_m(x) - G_n(x)| W(E_{m+n}(x)) \\
\Delta_{m,n}^{+} &= \sup_{x\in\mathbb{R}} \{F_m(x) - G_n(x)\} W(E_{m+n}(x)) \\
\Delta_{m,n}^{-} &= \sup_{x\in\mathbb{R}} \{G_n(x) - F_m(x)\} W(E_{m+n}(x)),
\end{aligned} \tag{1}$$

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 $\boldsymbol{Z}_{m+n} := (X_1, \dots, X_m, Y_1, \dots Y_n)$, and $W(t)$ is a non-negative weight function, defined on $t \in [0, 1]$.

The choice of weight function in (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 (1). For instance, if $W(t) \equiv 1$, the latter test, denoted as $\Delta_{m,n}^{0}$ and referred to as the unweighted KS test is given as:

$$\Delta_{m,n}^{0} = \sup_{x} |F_m(x) - G_n(x)|. \tag{2}$$

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 [26]), is defined as:

$$\Delta^{1/2}_{m,n} = \sup_x \frac{|F_m(x) - G_n(x)|}{\sqrt{E_{m+n}(x)(1 - E_{m+n}(x))}}, \tag{3}$$

It is not difficult to see that the weight functions from (2) and (3) generalize to (see [27]):

$$W(t) = 1/[t(1-t)]^{\nu}, \tag{4}$$

where $0 \le \nu \le 1$. The latter defines a family of weighted KS tests, $\Delta^{\nu}_{m,n}$ $(0 \le \nu \le 1)$, 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 [27] 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 [25] 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 part C of the supplementary material).

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 (10) and its further enhancement given in formula (11). Note that a recurrence method similar to that given in formula (10) but for the unweighted KS test has been independently considered by Nikiforov [3], Hilton et al. [23], and Schröer and Trenkler [24], but without theoretical justification.

We also show that the formula (10) can be further simplified to the recurrence formula (11), which leads to an algorithm recently considered by Viehmann [4].

The latter is slower but more numerically stable and accurate, as demonstrated in Section 4. We have implemented both formulas (10) and (11) in the C++ functions `ks2sample_cpp`, `ks2sample_c_cpp`, the R function `KS2sample` and in Mathematica [28], 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, $a_1 < a_2 < \cdots < a_k$, in the ordered, pooled sample $(z_1 \le z_2 \le \ldots \le z_{m+n})$, where $k \le m+n$. 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

$$P = \mathbb{P}\left(D_{m,n} \geq q|\boldsymbol{Z}_{m+n}\right), \tag{5}$$

where $D_{m,n}$ is the permutation KS statistic, defined similarly as in (1), but with respect to 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 (5) and write $\mathbb{P}(D_{m,n} \geq q)$.

Let us note that, $D_{m,n}$ in (5) is defined on the space, $\boldsymbol{\Omega}$ of all possible pairs, $C = \binom{m+n}{m}$ of edfs $F_m(x, \omega)$ and $G_n(x, \omega)$, $\omega \in \boldsymbol{\Omega}$, that correspond to the pairs of samples $\tilde{\boldsymbol{X}}_m$ and $\tilde{\boldsymbol{Y}}_n$, randomly drawn from, $\boldsymbol{Z}_{m+n}$, 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 $G_n(x, \omega)$, $\omega \in \boldsymbol{\Omega}$. 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 B.1). The above-mentioned resampling procedure is a general permutation and randomization tests construction (See [29], Section 15.2). An alternative definition of $F_m(x, \omega)$ and $G_n(x, \omega)$ based on permutation empirical measures can also be found in Section 3.8.1, van der Vaart and Wellner [30].

Considering the conditional p-value P in (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.*

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

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.

Table 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	$\lvert i/m - j/n\rvert$	$\lvert i/m - j/n\rvert W((i+j)/(m+n))$

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 [31]. Let us also note that a similar asymptotic consistency result for the unweighted KS test is provided by van der Vaart and Wellner [30] 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 (5), we alternatively express it as,

$$P = 1 - \frac{N}{C}, \tag{6}$$

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 $R = \{(i,j) : 0 \le i \le m, \quad 0 \le j \le n\}$. Define a trajectory on R as a point sequence $\{i_l, j_l\}_{l=0}^{m+n}$ that satisfies the following conditions

$$\begin{aligned} &1. \quad (i_{l+1}, j_{l+1}) = (i_l + 1, j_l) \quad \text{or} \quad (i_{l+1}, j_{l+1}) = (i_l, j_l + 1) \quad (0 \le l \le m+n-1) \\ &2. \quad (i_0, j_0) = (0,0) \quad \text{and} \quad (i_{m+n}, j_{m+n}) = (m,n) \end{aligned} \tag{7}$$

A trajectory moves to the right from (i_l, j_l) to $(i_l + 1, j_l)$ if z_l, the lth 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 part A of the supplementary material, 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,*

$$S = R_1 \cup R_2 \cup \cdots \cup R_k, \tag{8}$$

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

$$\begin{aligned} &\min\{i : d(i,j) < q, i+j = T_{l-1}\} \le i \le \max\{i : d(i,j) < q, i+j = T_l\}, \\ &\min\{j : d(i,j) < q, i+j = T_{l-1}\} \le j \le \max\{j : d(i,j) < q, i+j = T_l\} \end{aligned} \tag{9}$$

where $T_l = m_1 + \cdots + m_l$, $T_0 = 0$ and $d(i,j)$ is defined as shown in Table 1, depending on the particular definition of the KS test, $W(t)$ is the weight function defined in (1).

Proposition 2.2 is a generalization of formula (5) in [3]. 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 (8) 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 the proof of Proposition 2.2 provided in part A of the supplementary material.

Following Proposition 2.2, the subset, S, defined in (8) is then used to calculate the number, N, (c.f (6)), 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

$$B_S(i,j) = \mathbb{1}_S(i,j)[B_S(i,j-1) + B_S(i-1,j)] \quad \text{for } (i \geq 1, j \geq 1), \tag{10}$$

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 (10), 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 illustrating this algorithm in part B of the supplementary material.

Computing the p-value based on the recursion formula (10) can be further optimized to improve accuracy and numerical stability, in the spirit of [4]. 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 (10) as the recursion formula:

$$J_S(i,j) = 1 + \mathbb{1}_S(i,j)[\frac{i}{i+j}J_S(i-1,j) + \frac{i}{i+j}J_S(i,j-1) - 1] \quad \text{for } (i \geq 1, j \geq 1) \tag{11}$$

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 (11), 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 involves computing only 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.

2.3. 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,

$$\mathbb{P}\left\{\sqrt{\frac{nm}{n+m}}\Delta^0_{m,n} \leq x\right\} = 1 - 2\sum_{k=1}^{\infty}(-1)^{k-1}\,\mathrm{e}^{-2k^2x^2} \tag{12}$$

When the underlying null distributions $F = G$ are purely discrete, Eplett [32] 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 2.3 is based on several results. First, we recall the results of [30], 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 [5] for the asymptotic distribution of the unweighted one-sample KS test, assuming arbitrary null distribution, can be extended also to the two-sample case (see Theorem 2.3 and its proof in part A of the supplementary material).

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 $\boldsymbol{Z}_{m+n} = (X_1, \dots, X_m, Y_1, \dots, Y_n)$ in the unweighted KS test $D_{m,n}$ behaves as if it directly comes from the pooled distribution E.[2] 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 [5] for F in the one-sample case. More precisely, assume that E has finite number of jumps Λ, occurring at points x_l, $l = 1, \dots, \Lambda$, 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 $\nu_1, \nu_2, \dots$ the number of increasing segments appearing consecutively, and by $\omega_1, \omega_2, \dots$ 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

$$
\begin{aligned}
&\{x_1, \dots, x_{\nu_1}, x_{\nu_1+1}, \dots x_{\nu_1+\omega_1}, x_{\nu_1+\omega_1+1}, \dots, x_{\nu_1+\omega_1+\nu_2}, x_{\nu_1+\omega_1+\nu_2+1}, \dots, \\
&\quad x_{\nu_1+\omega_1+\nu_2+\omega_2}, \dots, x_{\nu_1+\omega_1+\cdots+\omega_{p-1}+1}, \dots, x_{\nu_1+\omega_1+\cdots+\omega_{p-1}+\nu_p}, \\
&\quad x_{\nu_1+\omega_1+\cdots+\omega_{p-1}+\nu_p+1}, \dots x_{\nu_1+\omega_1+\cdots+\omega_{p-1}+\nu_p+\omega_p}\} \qquad (13)
\end{aligned}
$$

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

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

Theorem 2.3: *Given the realization of* $\boldsymbol{X}_m$ *and* $\boldsymbol{Y}_n$ *coming from F and G respectively, denote by* $\Psi(x)$ *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* (13), *under the null hypothesis, when* $f_{2\Lambda} = f_{2\Lambda+1}$,

$$
\Psi(x) = \sum_{j_1=-\infty}^{\infty} \cdots \sum_{j_p=-\infty}^{\infty} \left((-1)^{j_1+\cdots+j_p}\right) c \int_{-x}^{x} \cdots \int_{-x}^{x} \mathrm{e}^{\alpha}\, \mathrm{d}z_1 \cdots \mathrm{d}z_{2\nu_p+w_p-1}, \qquad (14)
$$

where

$$c = \prod_{i=1}^{p}\left(\prod_{l=1}^{v_i}(f_{2(v_{i-1}+w_{i-1}+l)-1} - f_{2(v_{i-1}+w_{i-1}+l)-2})^{-1/2}\right.$$
$$\left.(f_{2(v_{i-1}+w_{i-1}+l)} - f_{2(v_{i-1}+w_{i-1}+l)-1})^{-1/2}\right)$$
$$\times\left(\prod_{l=1}^{\omega_i}\left(f_{2(v_i+w_{i-1}+l)} - f_{2(v_i+w_{i-1}+l)-1}\right)^{-1/2}\right)(2\pi)^{-\frac{2v_p+w_p-1}{2}}, \tag{15}$$

and

$$\alpha = -\frac{1}{2}\sum_{i=1}^{p}\left\{\sum_{l=1}^{v_i}\left[\frac{(z_{2(v_{i-1}+l)+w_{i-1}} - z_{2(v_{i-1}+l)+w_{i-1}-1})^2}{f_{2(v_{i-1}+w_{i-1}+l)} - f_{2(v_{i-1}+w_{i-1}+l)-1}}\right.\right.$$
$$\left.+\frac{(z_{2(v_{i-1}+l)+w_{i-1}-1} - (-1)^{j_{(v_{i-1}+l)}}z_{2(v_{i-1}+l)+w_{i-1}-2} - 2xj_{(v_{i-1}+l)})^2}{f_{2(v_{i-1}+w_{i-1}+l)-1} - f_{2(v_{i-1}+w_{i-1}+l)-2}}\right]$$
$$\left.+\sum_{l=1}^{\omega_i}\left[\frac{(z_{2v_i+w_{i-1}+l} - z_{2v_i+w_{i-1}+l-1})^2}{f_{2(v_i+w_{i-1}+l)} - f_{2(v_i+w_{i-1}+l)-1}}\right]\right\}, \tag{16}$$

with $v_0 = \omega_0 = 0; v_0 = w_0 = 0; v_i = \sum_{k=1}^{i} v_k; w_i = \sum_{k=1}^{i}\omega_k, v_p + w_p = \Lambda,$ *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* α' *for* α *in* (14) *to compute* $\Psi(x)$*, where*

$$c' = c(f_{2\Lambda+1} - f_{2\Lambda})^{-1/2}(2\pi)^{-1/2} \quad \text{and} \quad \alpha' = \alpha + \frac{(-(-1)^{j_{v_p+1}}z_{2v_p+w_p} - 2xj_{v_p+1})^2}{f_{2\Lambda+1} - f_{2\Lambda}} \tag{17}$$

Corollary 2.4: *When* $E(x)$ *is purely discrete with jumps* Λ*, the limiting distribution* $\Psi(x)$ *in (14) becomes:*

$$\Psi(x) = (2\pi)^{-\frac{\Lambda-1}{2}}\prod_{l=1}^{\Lambda}\left(f_{2l} - f_{2l-1}\right)^{-\frac{1}{2}}\int_{-x}^{x}\cdots\int_{-x}^{x}$$
$$\times\exp\left[-\frac{1}{2}\left(\sum_{l=1}^{\Lambda}\frac{(z_l - z_{l-1})^2}{f_{2l} - f_{2l-1}}\right)\right]\mathrm{d}z_1\cdots\mathrm{d}z_{\Lambda-1}. \tag{18}$$

Note that in view of Theorem 2.1, (18) also applies to the limiting distribution of $\Delta^0_{m,n}$, it is a closed-form explicit alternative to the result of [32], who only states convergence to a Gaussian process.

Remark 2.5: 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 2.3 and Corollary 2.4 can be applied to compute p-values of $D_{m,n}$, for large m and n. This is illustrated in Example 3.2, Section 3. In addition, for a fixed x, since $\frac{\partial}{\partial f_i}\Psi(x)$ 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 $|\Psi(x) - \hat{\Psi}(x)|$ introduced by estimation is also of order $O_p((m+n)^{-\frac{1}{2}})$, where $\hat{\Psi}(x)$ is computed by (14) based on $\hat{f}_l$, $l = 1, \ldots, 2\Lambda$.

2.4. The two-sample Kuiper test

The (unweighted) two sample Kuiper statistic [33] 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

$$\varsigma_{m,n} = \sup_{x\in\mathbb{R}}[F_m(x) - G_n(x)] - \inf_{x\in\mathbb{R}}[F_m(x) - G_n(x)], \tag{19}$$

and can be alternatively expressed as

$$\varsigma_{m,n} \equiv \Delta^{0+}_{m,n} + \Delta^{0-}_{m,n}, \tag{20}$$

where $\Delta^{0+}_{m,n}$ and $\Delta^{0-}_{m,n}$ are the one-sided two-sample KS tests defined in (1). As shown by Kuiper [33], the Kuiper statistic is invariant to cyclic transformations, in particular, invariant under all shifts and parametrizations on the circle (as shown in Example 3.1). 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 [34]. 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 [35], when the first sample from a Normal and the second from a mixture of normal [36] and Wyłomańska et al. [37]. Lemeshko et al. [38] 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.

2.5. Computing exact P values of the Kuiper test

As highlighted in (20), 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 (10) and (11) for computing the p-value of the KS test. Note that this step is also a generalization of the approaches by Maag and Stephens [21] and Hirakawa [22], 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 (5), the p-value for the permutation Kuiper test is defined as the probability

$$P' = \mathbb{P}\left(V_{m,n} \geq q|\boldsymbol{Z}_{m+n}\right), \tag{21}$$

where $V_{m,n}$ is the permutation Kuiper statistic, defined as in (19), on the space $\boldsymbol{\Omega}$, 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 2.6: *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 (6), the p-value in (21) 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 $\mathbf{Z}_{m+n}$, 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 2.7: *The number N', of pairs $F_m(x,\omega)$ and $G_n(x,\omega)$ for which $V_{m,n} < q$ is calculated as:*

$$N' = \begin{cases} \sum_{i=1}^{\lceil qr \rceil} N\left(\frac{i}{r}, \frac{\lceil qr \rceil - i + 1}{r}\right) - \sum_{i=1}^{\lceil qr \rceil - 1} N(\frac{i}{r}, \frac{\lceil qr \rceil - i}{r}) & q > 1/r \\ N\left(\frac{1}{r}, \frac{1}{r}\right) & q \le 1/r \end{cases} \tag{22}$$

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 $\mathbf{Z}_{m+n}$, 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 $\boldsymbol{\Omega}$, 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

$$S(a,b) = R_1(a,b) \cup R_2(a,b) \cup \cdots \cup R_k(a,b), \tag{23}$$

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

$$\begin{aligned} &\min\{i : (i/m - j/n) < a, i+j = T_{l-1}\} \le i \le \max\{i : (j/n - i/m) < b, i+j = T_l\}, \\ &\min\{j : (j/n - i/m) < b, i+j = T_{l-1}\} \le j \le \max\{j : (i/m - j/n) < a, i+j = T_l\} \end{aligned} \tag{24}$$

Remark 2.8: When $1 < q < 2$, the alternative for calculating the number N' is

$$N' = \sum_{i=\lceil qr \rceil - r}^{r+1} N\left(\frac{i}{r}, \frac{\lceil qr \rceil - i + 1}{r}\right) - \sum_{i=\lceil qr \rceil - r}^{r} N\left(\frac{i}{r}, \frac{\lceil qr \rceil - i}{r}\right), \tag{25}$$

which is a simplification of (22), 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 (22) via the number $B_S(m,n)$ in (10) 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 (22). 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 [4]. 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, (22) could be rewritten as a formula of the p-value P'

$$P' = \begin{cases} \sum_{i=1}^{\lceil qr \rceil} P\left(\frac{i}{r}, \frac{\lceil qr \rceil - i + 1}{r}\right) - \sum_{i=1}^{\lceil qr \rceil - 1} P\left(\frac{i}{r}, \frac{\lceil qr \rceil - i}{r}\right) & q > 1/r \\ P\left(\frac{1}{r}, \frac{1}{r}\right) & q \leq 1/r \end{cases} \tag{26}$$

Note that each value $P(a,b)$ in (26) can be directly computed from (11) using the number $J_S(m,n)$, substituting the set S with $S(a,b)$, defined in (23). 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 $P(\frac{i}{r}, \frac{\lceil qr \rceil - i}{r})$.

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.

3. Numerical implementation and examples

As mentioned in Sections 2.2 and 2.5, we have implemented the proposed efficient and exact calculation of p-values in C++, R and Mathematica [28]. Following (10) and (22), the Mathematica implementation (see footnote 1) computes the KS and Kuiper p-values to arbitrary precision, albeit at a high computational cost that grows with the sample sizes, as detailed in part D of the supplementary material. We use the latter p-values as the true values for comparison purposes. We have implemented (11) and (10) correspondingly in the C++ functions `ks2sample_cpp` and `ks2sample_c_cpp`, and (26) and (22) 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 part B of the supplementary material.

As illustrated in Section 4 and part D of the supplementary material, 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 (11) 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 4).

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

Example 3.1: Assume we have two discrete samples `sample3` and `sample4` observed on a circle, containing angular observations in degrees or radians. The `sample3` and `sample4`

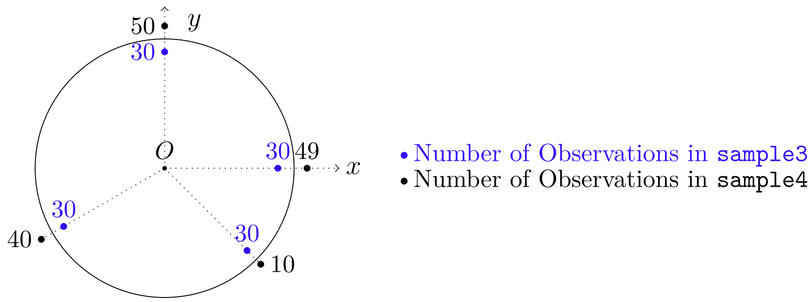

Figure 1. The position of observations in `sample3` and `sample4`.

contain 120 and 149 observations respectively, with all the observations equal to one of four angles, as shown in Figure 1.

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 = sample3, 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 (14) and the exact p-value when the data samples contain ties.

Example 3.2: Let us assume that we have a realization of the order statistic $\boldsymbol{Z}_{m+n}$ as follows:

$$z_1 = z_2 = \cdots = z_{1/2(m+n)} < z_{1/2(m+n)+1} < \cdots < z_{4/5(m+n)} = z_{4/5(m+n)+1} = \cdots = z_{m+n} \tag{27}$$

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 [5]. In order to apply (14) to compute the asymptotic p-value, one has to extract the jump structure of E from (27), 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 (14), the corresponding asymptotic distribution of the KS test is

$$\hat{\Psi}(x) = \left[\prod_{l=1}^{3}(\hat{f}_{l+1} - \hat{f}_l)\right]^{-\frac{1}{2}} \sum_{j=-\infty}^{\infty} \frac{(-1)^j}{2\pi} \int_{-x}^{x}\int_{-x}^{x} \mathrm{e}^{\alpha}\, \mathrm{d}z_1 z_2, \tag{28}$$

where

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

To compare the asymptotic p-value with the exact p-value, we need to implement $1 - \hat{\Psi}(x)$ in (28) 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. 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 $m + n \geq 90,000$.

4. Comparison with existing statistical software

The purpose of this section is to summarize the numerical comparison of the methods for computing KS and Kuiper p-values presented in Section 2 and implemented in the R functions `KS2sample` and `Kuiper2sample`, against nineteen implementations of the two-sample KS test and six implementations of the Kuiper test available in mainstream statistical software; the details of these are given in Sections D.1 and D.2 and part D of the supplementary material. We discuss their functionalities and select representative implementations of each test for comparison with our R functions. To assess accuracy, we use our Mathematica implementation (see footnote 1) to compute the true p-values and estimate

Table 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 $1 - \hat{\Psi}(1)$ computed by implementing (28) in Mathematica.

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

Note: Numbers in () are relative run times to the Mathematica implementation of (28).

the relative error of every competing implementation; for speed, we report the corresponding CPU run times on a machine with a 3.00 GHz Core i7-9700 processor and 32 GB RAM running Windows 10.

Detailed analysis of the results of this comparison can be found in part D of the supplementary material. In summary, we have shown in Table 5 that the R function `KS2sample` is very fast and accurate for both the unweighted and weighted KS tests and overall outperforms all competitors. The only viable alternative is 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 $m + n \leq 100,000$.

As for the function `Kuiper2sample` as we show in Section D.2 it is superior to all competitors in terms of accuracy, being the only function that gives the Kuiper p value with at least 8 correct digits.

5. Conclusions and discussion

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 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 [3], Hilton et al. [23], and Schröer and Trenkler [24], but without any theoretical justification.

In this work, we have provided its thorough theoretical justification, given by Theorem 2.1 its formal description (see Sections 2.1 and 2.2) and related proofs (see part A of the supplementary material) that are missing from works of these authors (see e.g. [3]), followed by some enlightening examples (see part B). We have generalized Nikiforov's method to accommodate arbitrary weight functions. In Sections 2.2 and 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 [21,22]. Following [4], we have further enhanced the algorithm in order to improve the accuracy of computing KS and Kuiper p-values. We have implemented both versions of this algorithm as the R functions `KS2sample` and `Kuiper2sample`, illustrating their use in Section 3, and also provide Mathematica code for arbitrary-accuracy computation (see its availability at footnote 1).

In Section 2.3 we have derived a closed form formula (c.f. (14), Theorem 2.3) 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 (14) and Theorem 2.3 represent a novel contribution of theoretical and numerical importance. We have demonstrated (see Example 3.2 and Table 2) that the asymptotic formula (14) 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 paper 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 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 Tables 5 and 8, part D of the supplementary material). For the Kuiper test, `Kuiper2sample` is the only function that provides exact p-values with at least 8 correct digits (see Table 10, part D).

Last but not least, 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 C.1 and C.2, part C of the supplementary material, 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.

As direction of future research one could consider generalizing the approach to the case of two-sample Cramér–von Mises and Anderson–Darling tests. The theoretical justification of Theorem 2.1 directly extends to these tests, as its proof relies only on the weak convergence of the permutation empirical process, which holds for any continuous functional of the empirical process. However, the recurrence formula of Section 2 does not directly extend to these tests, since they are defined as weighted integrals rather than suprema of the empirical distribution functions and do not admit a trajectory-counting formulation on the integer-valued grid R. Extending the recurrence formula of Section 2 to compute exact permutation p-values of the two-dimensional two-sample KS test of [39,40] is a non-trivial open problem. Unlike the one-dimensional case, there is no canonical total ordering of points in $\mathbb{R}^2$, and the two-dimensional KS statistic at each node depends on the full spatial arrangement of all previously assigned observations rather than on the grid coordinates alone, so the Markov property of the lattice path breaks down and the recurrence formula (10) does not directly generalize. Developing an efficient exact algorithm for this case is a direction of future research.

Notes

1. The Mathematica code is available at https://github.com/fakecloudsjy/KS2sample-Kuiper2sample.
2. This consideration relies on Theorem 2.1, and it fails to hold in a more general case. See more discussion in [31].

Author contributions

CRediT: **Dimitrina S. Dimitrova:** Methodology, Software, Supervision, Validation, Writing – original draft, Writing – review & editing; **Yun Jia:** Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing; **Vladimir K. Kaishev:** Conceptualization, Formal analysis, Investigation, Methodology, Supervision, Validation, Writing – original draft, Writing – review & editing

Disclosure statement

No potential conflict of interest was reported by the author(s).

References

[1] Borrajo M, González-Manteiga W, Martínez-Miranda M. Goodness-of-fit test for point processes first-order intensity. Comput Stat Data Anal. 2024;194:107929. doi: 10.1016/j.csda.2024.107929. Available from: https://www.sciencedirect.com/science/article/pii/S0167947324000136
[2] Meintanis S, Milošević B, Jiménez–Gamero M. Goodness–of–fit tests based on the min–characteristic function. Comput Stat Data Anal. 2024;197:107988. doi: 10.1016/j.csda.2024.107988. Available from: https://www.sciencedirect.com/science/article/pii/S0167947324000720
[3] Nikiforov AM. Algorithm as 288: exact Smirnov two-sample tests for arbitrary distributions. J R Stat Soc Ser C (Appl Stat). 1994;43(1):265–270.
[4] Viehmann T. Numerically more stable computation of the p-values for the two-sample kolmogorov-smirnov test; 2021.
[5] Dimitrova DS, Kaishev VK, Tan S. Computing the Kolmogorov-Smirnov distribution when the underlying CDF is purely discrete, mixed, or continuous. J Stat Softw. 2020;95(10):1–42. doi: 10.18637/jss.v095.i10
[6] Moscovich A. Fast calculation of p-values for one-sided kolmogorov-smirnov type statistics. Comput Stat Data Anal. 2023;185:107769. doi: 10.1016/j.csda.2023.107769. Available from: https://www.sciencedirect.com/science/article/pii/S0167947323000804
[7] Paltani S. Searching for periods in x-ray observations using Kuiper's test-application to the rosat pspc archive. Astron Astrophys. 2004;420(2):789–797. doi: 10.1051/0004-6361:20034220
[8] Feigelson E, Babu GJ. Beware the kolmogorov-smirnov test!; 2020. [cited 2024 Nov 1]. Available from: https://asaip.psu.edu/articles/beware-the-kolmogorov-smirnov-test/

[9] Mousavi M, Davulcu H, Ahmadi M, et al. Effective messaging on social media: what makes online content go viral? In: Proceedings of the ACM Web Conference 2022. New York (NY), USA: Association for Computing Machinery; 2022.
[10] Arsioli B, Dedin P. Machine learning applied to multifrequency data in astrophysics: Blazar classification. Mon Not R Astron Soc. 2020;498(2):1750–1764. doi: 10.1093/mnras/staa2449
[11] Zhang W, Huang J, Zhang T, et al. A risk-based stochastic model for supporting resources allocation of agricultural water-energy-food system under uncertainty. J Hydrol. 2022;610:127864. doi: 10.1016/j.jhydrol.2022.127864
[12] Yung C, Çolak G, Wang W. Cycles in the IPO market. J Financ Econ. 2008;89(1):192–208. doi: 10.1016/j.jfineco.2007.06.007
[13] Dudbridge F. A note on permutation tests in multistage association scans. Am J Hum Genet. 2006;78(6):1094–1095. doi: 10.1086/504527
[14] Mouli SC, Teixeira L, Neville J, et al. Deep lifetime clustering; 2019.
[15] Lin LC-C, Hu C-P, Takata J, et al. Investigation of the timing and spectral properties of an ultraluminous x-ray pulsar ngc 7793 p13. Astrophys J. 2022;924(2):65. doi: 10.3847/1538-4357/ac32b9
[16] Chhatbar PY, Francis JT. Towards a naturalistic brain-machine interface: hybrid torque and position control allows generalization to novel dynamics. PLoS ONE. 2013;8(1):1–9. doi: 10.1371/journal.pone.0052286
[17] Kossobokov VG, Panza GF. A myth of preferred days of strong earthquakes? Seismol Res Lett. 2020;91(2A):948–955. doi: 10.1785/0220190157
[18] Hodges JL. The significance probability of the Smirnov two-sample test. Ark Mat. 1958;3(5):469–486. doi: 10.1007/BF02589501
[19] Kim PJ. On the exact and approximate sampling distribution of the two sample Kolmogorov-Smirnov criterion d_{mn}, m $\leq$ n. J Am Stat Assoc. 1969;64(328):1625–1637.
[20] Kim P, Jennrich RI. Tables of the exact sampling distribution of the two-sample kolmogorov-smirnov criterion, d_{mn}, m $\leq$ n. In: Selected tables in mathematical statistics. Vol. 1. Providence (RI): American Mathematical Society; 1973. p. 80–129.
[21] Maag UR, Stephens MA. The V_{NM} two-sample test. Ann Math Stat. 1968;39(3):923–935. doi: 10.1214/aoms/1177698324
[22] Hirakawa K. The two-sample Kuiper test. TRU Math. 1973;9:99–118.
[23] Hilton JF, Mehta CR, Patel NR. An algorithm for conducting exact smirnov tests. Comput Stat Data Anal. 1994;17(4):351–361. doi: 10.1016/0167-9473(94)90017-5
[24] Schröer G, Trenkler D. Exact and randomization distributions of Kolmogorov-Smirnov tests two or three samples. Comput Stat Data Anal. 1995;20(2):185–202. doi: 10.1016/0167-9473(94)00040-P
[25] Büning H. Kolmogorov-smirnov- and cramér-von mises type two-sample tests with various weight functions. Commun Stat B: Simul Comput. 2001;30(4):847–865. doi: 10.1081/SAC-100107784
[26] Canner PL. A simulation study of one- and two-sample Kolmogorov–Smirnov statistics with a particular weight function. J Am Stat Assoc. 1975;70(349):209–211. doi: 10.1080/01621459.1975.10480290
[27] Finner H, Gontscharuk V. Two-sample Kolmogorov–Smirnov-type tests revisited: old and new tests in terms of local levels. Ann Stat. 2018;46(6A):3014–3037. doi: 10.1214/17-AOS1647
[28] Wolfram Research, Inc. Mathematica, Version 14.0. Champaign, Illinois; 2024.
[29] Lehmann EL, Romano JP. Testing statistical hypotheses (3rd ed.). New York: Springer; 2005.
[30] van der Vaart AW, Wellner JA. Weak convergence and empirical processes. 2nd ed. Cham: Springer; 2023. (Springer series in statistics).
[31] Chung E, Romano JP. Exact and asymptotically robust permutation tests. Ann Stat. 2013;41(2):484–507. doi: 10.1214/13-AOS1090
[32] Eplett WJR. The distributions of Smirnov type two-sample rank tests for discontinuous distribution functions. J R Stat Soc: Ser B (Methodol). 1982;44(3):361–369. doi: 10.1111/j.2517-6161.1982.tb01214.x

[33] Kuiper NH. Tests concerning random points on a circle. Proc K Ned Akad Wet, Ser A. 1960;63:38–47.
[34] Press WH, Teukolsky SA, Vetterling WT, et al. Numerical recipes: the art of scientific computing (3rd ed.). New York: Cambridge University Press; 2007.
[35] Foutz RV, Birch JB. Tests for the two-sample problem based on empirical probability measures. Commun Stat – Theory Methods. 1982;11(16):1839–1853. doi: 10.1080/03610928208828354
[36] Dowd C. A new ecdf two-sample test statistic; 2020. Available from: arXiv preprint arXiv:2007.01360.
[37] Wyłoma#324;ska A, Iskander DR, Burnecki K. Omnibus test for normality based on the edgeworth expansion. PLoS ONE. 2020;15(6):1–36.
[38] Lemeshko B, Gorbunova A, Lemeshko S, et al. Solving problems of using some nonparametric goodness-of-fit tests. Optoelectron Instrum Data Process. 2014;50:21–35. doi: 10.3103/S875669901401004X
[39] Fasano G, Franceschini A. A multidimensional version of the Kolmogorov–Smirnov test. Mon Not R Astron Soc. 1987;225(1):155–170. doi: 10.1093/mnras/225.1.155
[40] Peacock JA. Two-dimensional goodness-of-fit testing in astronomy. Mon Not R Astron Soc. 1983;202(3):615–627. doi: 10.1093/mnras/202.3.615