# Statistical tests for whether a given set of independent, identically distributed draws comes from a specified probability density

**Mark Tygert** *

*Courant Institute of Mathematical Sciences, NYU, New York, NY 10012

Submitted to Proceedings of the National Academy of Sciences of the United States of America

**We discuss several tests for whether a given set of independent and identically distributed (i.i.d.) draws does not come from a specified probability density function. The most commonly used are Kolmogorov-Smirnov tests, particularly Kuiper's variant, which focus on discrepancies between the cumulative distribution function for the specified probability density and the empirical cumulative distribution function for the given set of i.i.d. draws. Unfortunately, variations in the probability density function often get smoothed over in the cumulative distribution function, making it difficult to detect discrepancies in regions where the probability density is small in comparison with its values in surrounding regions. We discuss tests without this deficiency, complementing the classical methods. The tests of the present paper are based on the plain fact that it is unlikely to draw a random number whose probability is small, provided that the draw is taken from the same distribution used in calculating the probability (thus, if we draw a random number whose probability is small, then we can be confident that we did not draw the number from the same distribution used in calculating the probability).**

Kolmogorov-Smirnov | nonparametric | goodness-of-fit | outlier | distribution function | nonincreasing rearrangement

A basic task in statistics is to ascertain whether a given set of independent and identically distributed (i.i.d.) draws $X_1$, $X_2$, ..., $X_{n-1}$, $X_n$ does not come from a distribution with a specified probability density function $p$ (the null hypothesis is that $X_1$, $X_2$, ..., $X_{n-1}$, $X_n$ do in fact come from the specified $p$). In the present paper, we consider the case when $X_1$, $X_2$, ..., $X_{n-1}$, $X_n$ are real valued. In this case, the most commonly used approach is due to Kolmogorov and Smirnov (with a popular modification by Kuiper); see, for example, Sections 14.3.3 and 14.3.4 of [1], [2], [3], or *Test Statistics* below.

The Kolmogorov-Smirnov approach considers the size of the discrepancy between the cumulative distribution function for $p$ and the empirical cumulative distribution function defined by $X_1$, $X_2$, ..., $X_{n-1}$, $X_n$ (see, for example, *Notation* and *Test Statistics* below for definitions of cumulative distribution functions and empirical cumulative distribution functions). If the i.i.d. draws $X_1$, $X_2$, ..., $X_{n-1}$, $X_n$ used to form the empirical cumulative distribution function are taken from the probability density function $p$ used in the Kolmogorov-Smirnov test, then the discrepancy is small. Thus, if the discrepancy is large, then we can be confident that $X_1$, $X_2$, ..., $X_{n-1}$, $X_n$ do not come from a distribution with probability density function $p$.

However, the size of the discrepancy between the cumulative distribution function for $p$ and the empirical cumulative distribution function constructed from the i.i.d. draws $X_1$, $X_2$, ..., $X_{n-1}$, $X_n$ does not always signal that $X_1$, $X_2$, ..., $X_{n-1}$, $X_n$ do not arise from a distribution with the specified probability density function $p$, even when $X_1$, $X_2$, ..., $X_{n-1}$, $X_n$ do not in fact arise from $p$. In some cases, $n$ has to be absurdly large for the discrepancy to be significant. It is easy to see why:

The cumulative distribution function is an indefinite integral of the probability density function $p$. Therefore, the cumulative distribution function is a smoothed version of the probability density function; focusing on the cumulative distribution function rather than $p$ itself makes it harder to detect discrepancies in regions where $p$ is small in comparison with its values in surrounding regions. For example, consider the probability density function $p$ depicted in Figure 1 below (a "tent" with a narrow triangle removed at its apex) and the probability density function $q$ depicted in Figure 2 below (nearly the same "tent," but with the narrow triangle intact, not removed). The cumulative distribution functions for $p$ and $q$ are very similar, so tests of the classical Kolmogorov-Smirnov type have trouble signaling that i.i.d. draws taken from $q$ are actually not taken from $p$. Section 14.3.4 of [1] highlights this problem and a strategy for its solution, hence motivating us to write the present article.

We propose to supplement tests of the classical Kolmogorov-Smirnov type with tests for whether any of the values $p(X_1)$, $p(X_2)$, ..., $p(X_{n-1})$, $p(X_n)$ is small. If any of these values is small, then we can be confident that the i.i.d. draws $X_1$, $X_2$, ..., $X_{n-1}$, $X_n$ did not arise from the probability density function $p$. Theorem 3 below formalizes the notion of any of $p(X_1)$, $p(X_2)$, ..., $p(X_{n-1})$, $p(X_n)$ being small. We also propose another complementary test, which amounts to using the Kolmogorov-Smirnov approach after "rearranging" the probability density function $p$ so that it is nondecreasing on the shortest interval outside which it vanishes (see Remark 2 and Eq. **4** below).

For descriptions of other generalizations of and alternatives to the Kolmogorov-Smirnov approach (concerning issues distinct from those treated in the present paper), see, for example, Sections 14.3.3 and 14.3.4 of [1], [2], [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], and their compilations of references. For a more general approach, based on customizing statistical tests for problem-specific families of alternative hypotheses, see [14]. Below, we compare the test statistics of the present article with one of the most commonly used test statistics of the Kolmogorov-Smirnov type, namely Kuiper's (see, for example, [2], [3], or *Test Statistics* below). We recommend using the test statistics of the present paper in conjunction with the Kuiper statistic, to be conservative, as all these statistics complement each other, helping compensate for their inevitable deficiencies.

There are at least two canonical applications. First, the tests of the present article can be suitable for checking for malfunctions with and bugs in computer codes that are supposed to generate pseudorandom i.i.d. draws from specified probability density functions (especially the complicated ones encountered frequently in practice). Good software engineering requires such independent tests for help-

**Reserved for Publication Footnotes**

ing validate that computer codes produce correct results (of course, such validations do not obviate careful, structured programming, but are instead complementary). Second, many theories from physics and physical chemistry predict (often *a priori*) the probability density functions from which experiments are supposed to be taking i.i.d. draws. The tests of the present paper can be suitable for ruling out erroneous theories of this type, on the basis of experimental data. Moreover, there are undoubtedly many other potential applications, in addition to these two.

For definitions of the notation used throughout this article, see *Notation* below. *Test Statistics* introduces several statistical tests. *Numerical Examples* illustrates the power of the statistical tests. *Conclusions and Generalizations* draws some conclusions and proposes directions for further work.

**Remark 1.** *All tests used in the present paper do not require any intervention by the user of suitable software implementations. The tests are not a panacea; all such tests have the drawbacks discussed in [14]. See [14] for a much more flexible alternative, allowing the user to amend tests to be more powerful against user-specified parametric families of alternative hypotheses.*

## Notation

In this section, we set notation used throughout the present paper.

We use $\mathbf{P}$ to take the probability of an event. We say that $p$ is a probability density function to mean that $p$ is a (Lebesgue-measurable) function from $\mathbb{R}$ to $[0, \infty)$ such that the integral of $p$ over $\mathbb{R}$ is 1.

The cumulative distribution function $P$ for a probability density function $p$ is

$$P(x) = \int_{y \le x} p(y)\, dy \qquad \textbf{[1]}$$

for any real number $x$. If $X$ is a random variable distributed according to $p$, then $P(x)$ is just the probability that $X \le x$. Therefore, if $X$ is a random variable distributed according to $p$, then the cumulative distribution function $\mathcal{P}$ for $p(X)$ is

$$\mathcal{P}(x) = \int_{p(y) \le x} p(y)\, dy, \qquad \textbf{[2]}$$

the probability that $p(X) \le x$.

For reference, we summarize our (reasonably standard) notational conventions in Table 1.

**Remark 2.** *The "nonincreasing rearrangement" (or nondecreasing rearrangement) of a probability density function (see, for example, Section V.3 of [15]) clarifies the meaning of the distribution function $\mathcal{P}$ defined in Eq.* **2***. With $P$ defined in Eq.* **1** *and $\mathcal{P}$ defined in Eq.* **2***, $\mathcal{P}(p(x)) = P(x)$ for any real number $x$ in the shortest interval outside which the probability density function $p$ vanishes, as long as $p$ is increasing on that shortest interval.*

## Test Statistics

In this section, we introduce several statistical tests.

One test of whether i.i.d. draws $X_1$, $X_2$, …, $X_{n-1}$, $X_n$ do not arise from a specified probability density function $p$ is the Kolmogorov-Smirnov test (or Kuiper's often preferable variation). If $X$ is a random variable distributed according to $p$, then another test is to use the Kolmogorov-Smirnov or Kuiper test for the random variable $p(X)$, whose cumulative distribution function is $\mathcal{P}$ in Eq. **2**. The test statistic for the original Kuiper test is

$$U = \left( \sqrt{n} \sup_{-\infty < x < \infty} P(x) - \hat{P}(x) \right) - \left( \sqrt{n} \inf_{-\infty < x < \infty} P(x) - \hat{P}(x) \right), \qquad \textbf{[3]}$$

where $\hat{P}(x)$ is the empirical cumulative distribution function — the number of $k$ such that $X_k \le x$, divided by $n$. The test statistic for the Kuiper test for $p(X)$ is therefore

$$V = \left( \sqrt{n} \sup_{0 \le x < \infty} \mathcal{P}(x) - \hat{\mathcal{P}}(x) \right) - \left( \sqrt{n} \inf_{0 \le x < \infty} \mathcal{P}(x) - \hat{\mathcal{P}}(x) \right), \qquad \textbf{[4]}$$

where $\hat{\mathcal{P}}(x)$ is the number of $k$ such that $p(X_k) \le x$, divided by $n$. Remark 2 above and Remark 5 below provide some motivation for using $V$, beyond its being a natural variation on $U$.

The rationale for using statistics such as $U$ and $V$ is the following theorem, corollary, and the ensuing discussion (see, for example, Sections 14.3.3 and 14.3.4 of [1], [3], or [2] for proofs and details).

**Theorem 1.** *Suppose that $p$ is a probability density function, $X$ is a random variable distributed according to $p$, and $P$ is the cumulative distribution function for $X$ from Eq.* **1***. Then, the distribution of $P(X)$ is the uniform distribution over $[0, 1]$.*

**Corollary 2.** *Suppose that $p$ is a probability density function, $X$ is a random variable distributed according to $p$, and $\mathcal{P}$ is the cumulative distribution function for $p(X)$ from Eq.* **2***. Then, the cumulative distribution function of $\mathcal{P}(p(X))$ is less than or equal to the cumulative distribution function of the uniform distribution over $[0, 1]$. Moreover, the distribution of $\mathcal{P}(p(X))$ is the uniform distribution over $[0, 1]$ if $\mathcal{P}$ is a continuous function ($\mathcal{P}$ is a continuous function when, for every nonnegative real number $y$, the probability that $p(X) = y$ is 0).*

Theorem 1 generalizes to the fact that, if the i.i.d. draws $X_1$, $X_2$, …, $X_{n-1}$, $X_n$ arise from the probability density function $p$ involved in the definition of $U$ in Eq. **3**, then the distribution of $U$ does not depend on $p$; the distribution of $U$ is the same for any $p$. With high probability, $U$ is not much greater than 1 when the i.i.d. draws $X_1$, $X_2$, …, $X_{n-1}$, $X_n$ used in the definition of $U$ in Eq. **3** are taken from the distribution whose probability density function $p$ and cumulative distribution function $P$ are used in the definition of $U$. Therefore, if the statistic $U$ that we compute turns out to be substantially greater than 1, then we can have high confidence that the i.i.d. draws $X_1$, $X_2$, …, $X_{n-1}$, $X_n$ were not taken from the distribution whose probability density function $p$ and cumulative distribution function $P$ were used in the definition of $U$. Similarly, if $V$ defined in Eq. **4** turns out to be substantially greater than 1, then we can have high confidence that the i.i.d. draws $X_1$, $X_2$, …, $X_{n-1}$, $X_n$ were not taken from the distribution whose probability density function $p$ and distribution function $\mathcal{P}$ were used in the definition of $V$. For details, see, for example, Sections 14.3.3 and 14.3.4 of [1], [3], or [2].

A third test statistic is

$$W = n \min_{1 \le k \le n} \mathcal{P}(p(X_k)). \qquad \textbf{[5]}$$

The following theorem (which follows immediately from Corollary 2) and ensuing discussion characterize $W$ and its applications.

**Theorem 3.** *Suppose that $p$ is a probability density function, $n$ is a positive integer, $X_1$, $X_2$, …, $X_{n-1}$, $X_n$ are i.i.d. random variables each distributed according to $p$, $\mathcal{P}$ is the cumulative distribution function for $p(X_1)$ from Eq.* **2***, and $W$ is the random variable defined in Eq.* **5***. Then,*

$$\mathbf{P}\{W \le x\} \le 1 - \left(1 - \frac{x}{n}\right)^n \qquad \textbf{[6]}$$

*for any $x \in [0, n]$.*

**Table 1. Notational conventions**

| *mathematical object* | *typeface* | *example* |
|---|---|---|
| probability density function | italic lowercase | $p(x)$ |
| cumulative distribution func. in Eq. **1** | italic uppercase | $P(x)$ |
| distribution function defined in Eq. **2** | script uppercase | $\mathcal{P}(x)$ |
| taking the probability of an event | bold uppercase | $\mathbf{P}\{X \le x\}$ |

For any positive real number $\alpha < 1/2$, we define

$$x_\alpha = n - n(1-\alpha)^{1/n}; \qquad \textbf{[7]}$$

if $W \le x_\alpha$, then due to Eq. **6** we can have at least $[100(1-\alpha)]\%$ confidence that the i.i.d. draws $X_1, X_2, \ldots, X_{n-1}, X_n$ do not arise from $p$. It follows from Eq. **7** that

$$\alpha \le x_\alpha < -\ln(1-\alpha) = \alpha + \alpha^2/2 + \alpha^3/3 + \ldots < \alpha + \alpha^2, \qquad \textbf{[8]}$$

with $x_\alpha = \alpha$ for $n = 1$, and $\lim_{n\to\infty} x_\alpha = -\ln(1-\alpha)$. Therefore, if $W \le \alpha$, then we have at least $[100(1-\alpha)]\%$ confidence that the i.i.d. draws $X_1, X_2, \ldots, X_{n-1}, X_n$ do not arise from $p$. Taking $\alpha = .01$, for example, we have at least 99% confidence that the i.i.d. draws $X_1, X_2, \ldots, X_{n-1}, X_n$ do not arise from $p$, if $W \le .01$.

**Remark 3.** *If $W$ defined in Eq.* **5** *is at most* 1*, then we can have at least $[100(1-W)]\%$ confidence that the i.i.d. draws $X_1, X_2, \ldots, X_{n-1}, X_n$ do not arise from the probability density function $p$ used in Eq.* **5***.*

**Remark 4.** *Using $W$ defined in Eq.* **5** *along with the upper bound in Eq.* **6** *is optimal when the probability density function $p$ takes on only finitely many values, or when $p$ has the property that, for every nonnegative real number $y$, the probability is 0 that $p(X) = y$, where $X$ is a random variable distributed according to $p$. In both cases, the inequality in Eq.* **6** *becomes the equality*

$$\mathbf{P}\{W \le n\,\mathcal{P}(p(x))\} = 1 - \Big(1 - \mathcal{P}(p(x))\Big)^n \qquad \textbf{[9]}$$

*for any $x \in \mathbb{R}$.*

**Remark 5.** *When the statistic $W$ defined in Eq.* **5** *is not powerful enough to discriminate between two particular distributions, then a natural alternative is the average*

$$\tilde{W} = \frac{1}{n}\sum_{k=1}^{n} \mathcal{P}(p(X_k)). \qquad \textbf{[10]}$$

*The Kuiper test statistic $V$ defined in Eq.* **4** *is a more refined version of this alternative, and we recommend using $V$ instead of $\tilde{W}$, in conjunction with the use of $W$ and $U$ defined in Eq.* **3***. We could also consider more general averages of the form*

$$f\left(\frac{1}{n}\sum_{k=1}^{n} g\Big(\mathcal{P}(p(X_k))\Big)\right), \qquad \textbf{[11]}$$

*where $f$ and $g$ are functions; obvious candidates include $f(x) = \exp(x)$ and $g(x) = \ln(x)$, and $f(x) = 1 - x^{1/q}$ and $g(x) = (1-x)^q$, with $q \in (1, \infty)$.*

## Numerical Examples

In this section, we illustrate the effectiveness of the test statistics of the present paper via several numerical experiments. For each experiment, we compute the statistics $U$, $V$, and $W$ defined in Eqs. **3**, **4**, and **5** for two sets of i.i.d. draws, first for i.i.d. draws $X_1, X_2, \ldots, X_{n-1}, X_n$ taken from the distribution whose probability density function $p$, cumulative distribution function $P$, and distribution function $\mathcal{P}$ are used in the definitions of $U$, $V$, and $W$ in Eqs. **3**, **4**, and **5**, and second for i.i.d. draws $X_1, X_2, \ldots, X_{n-1}, X_n$ taken from a different distribution.

The test statistics $U$ and $V$ defined in Eqs. **3** and **4** are the same, except that $U$ concerns a random variable $X$ drawn from a probability density function $p$, while $V$ concerns $p(X)$. We can directly compare the values of $U$ and $V$ for various distributions in order to gauge their relative discriminative powers. Ideally, $U$ and $V$ should not be much greater than 1 when the i.i.d. draws $X_1, X_2, \ldots, X_{n-1}, X_n$ used in the definitions of $U$ and $V$ in Eqs. **3** and **4** are taken from the distribution whose probability density function $p$, cumulative distribution function $P$, and distribution function $\mathcal{P}$ are used in the definitions of $U$ and $V$; $U$ and $V$ should be substantially greater than 1 when the i.i.d. draws $X_1, X_2, \ldots, X_{n-1}, X_n$ are taken from a different distribution, to signal the difference between the common distribution of each of $X_1, X_2, \ldots, X_{n-1}, X_n$ and the distribution whose probability density function $p$, cumulative distribution function $P$, and distribution function $\mathcal{P}$ are used in the definitions of $U$ and $V$.

For details concerning the interpretation of and significance levels for the Kuiper test statistics $U$ and $V$ defined in Eqs. **3** and **4**, see Sections 14.3.3 and 14.3.4 of [1], [2], or [3]; both one- and two-tailed hypothesis tests are available, for any finite number $n$ of draws $X_1, X_2, \ldots, X_{n-1}, X_n$, and also in the limit of large $n$. In short, if $X_1, X_2, \ldots, X_{n-1}, X_n$ are i.i.d. random variables drawn according to a continuous cumulative distribution function $P$, then the complementary cumulative distribution function of $U$ defined in Eq. **3** for the same cumulative distribution function $P$ has an upper tail that decays nearly as fast as the complementary error function. Although the details are complicated (varying with $n$ and with the form — one-tailed or two-tailed — of the hypothesis test), the probability that $U$ is greater than 2 is at most 1% when $X_1, X_2, \ldots, X_{n-1}, X_n$ used in Eq. **3** are drawn according to the same cumulative distribution function $P$ as used in Eq. **3**.

As described in Remark 3, the interpretation of the test statistic $W$ defined in Eq. **5** is simple: If $W$ defined in Eq. **5** is at most 1, then we can have at least $[100(1-W)]\%$ confidence that the i.i.d. draws $X_1, X_2, \ldots, X_{n-1}, X_n$ do not arise from the probability density function $p$ used in Eq. **5**.

Tables 2–5 display numerical results for the examples described in the subsections below. The following list describes the headings of the tables:

- $n$ is the number of i.i.d. draws $X_1, X_2, \ldots, X_{n-1}, X_n$ taken to form the statistics $U$, $V$, and $W$ defined in Eqs. **3**, **4**, and **5**.
- $U_0$ is the statistic $U$ defined in Eq. **3**, with the $X_1, X_2, \ldots, X_{n-1}, X_n$ defining $\hat{P}$ in Eq. **3** drawn from a distribution with the same cumulative distribution function $P$ as used in Eq. **3**. Ideally, $U_0$ should be small, not much larger than 1.
- $U_1$ is the statistic $U$ defined in Eq. **3**, with the $X_1, X_2, \ldots, X_{n-1}, X_n$ defining $\hat{P}$ in Eq. **3** drawn from a distribution with a cumulative distribution function that is different from $P$ used in Eq. **3**. Ideally, $U_1$ should be large, substantially greater than 1, to signal the difference between the common distribution of each of $X_1, X_2, \ldots, X_{n-1}, X_n$ and the distribution with the cumulative distribution function $P$ used in Eq. **3**. The numbers in parentheses in the tables indicate the order of magnitude of the significance level for rejecting the null hypothesis, that is, for asserting that the draws $X_1, X_2, \ldots, X_{n-1}, X_n$ do not arise from $P$.
- $V_0$ is the statistic $V$ defined in Eq. **4**, with the $X_1, X_2, \ldots, X_{n-1}, X_n$ defining $\hat{\mathcal{P}}$ in Eq. **4** drawn from a distribution with the same probability density function $p$ used for $\hat{\mathcal{P}}$ and for $\mathcal{P}$ in Eq. **4**. Ideally, $V_0$ should be small, not much larger than 1.
- $V_1$ is the statistic $V$ defined in Eq. **4**, with the $X_1, X_2, \ldots, X_{n-1}, X_n$ defining $\hat{\mathcal{P}}$ in Eq. **4** drawn from a distribution that is different from the distribution with the probability density function $p$ used for $\hat{\mathcal{P}}$ and for $\mathcal{P}$ in Eq. **4**. Ideally, $V_1$ should be large, substantially greater than 1, to signal the difference between the common distribution of each of $X_1, X_2, \ldots, X_{n-1}, X_n$ and the distribution with the probability density function $p$ used for $\hat{\mathcal{P}}$ and for $\mathcal{P}$ in Eq. **4**. The numbers in parentheses in the tables indicate the order of magnitude of the significance level for rejecting the null hypothesis, that is, for asserting that the draws $X_1, X_2, \ldots, X_{n-1}, X_n$ do not arise from $p$. We used [2] to estimate the significance level; this estimate can be conservative for $V$.
- $W_0$ is the statistic $W$ defined in Eq. **5**, with the $X_1, X_2, \ldots, X_{n-1}, X_n$ in Eq. **5** drawn from a distribution with the same probability density function $p$ and distribution function $\mathcal{P}$ in Eq. **5**. Ideally, $W_0$ should not be much less than 1.

- $W_1$ is the statistic $W$ defined in Eq. **5**, with the $X_1$, $X_2$, …, $X_{n-1}$, $X_n$ in Eq. **5** drawn from a distribution that is different from the distribution with the probability density function $p$ used in Eq. **5** ($p$ is used both directly and for defining the distribution function $\mathcal{P}$ in Eq. **5**). Ideally, $W_1$ should be small, substantially less than 1, to signal the difference between the common distribution of each of $X_1$, $X_2$, …, $X_{n-1}$, $X_n$ and the distribution with the probability density function $p$ used in Eq. **5**. The value of $W_1$ itself is the significance level for rejecting the null hypothesis, that is, for asserting that the draws $X_1, X_2, \ldots, X_{n-1}, X_n$ do not arise from $p$.

**A sawtooth wave.** The probability density function $p$ for our first example is

$$p(x) = \begin{cases} 2\text{E}{-}3 \cdot (x - k), & x \in (k, k+1) \text{ for } k \in \{0, 1, \ldots, 999\} \\ 0, & \text{otherwise} \end{cases} \quad \textbf{[12]}$$

for any $x \in \mathbb{R}$.

We compute the statistics $U$, $V$, and $W$ defined in Eqs. **3**, **4**, and **5** for two sets of i.i.d. draws, first for i.i.d. draws distributed according to $p$ defined in Eq. **12**, and then for i.i.d. draws from the uniform distribution on $(0, 1000)$. Table 2 displays numerical results.

For this example, the classical Kuiper statistic $U$ is unable to signal that the draws from the uniform distribution do not arise from $p$ defined in Eq. **12** for $n \leq 10^7$, at least not nearly as well as the modified Kuiper statistic $V$, which signals the discrepancy with very high confidence for $n \geq 10^3$. The statistic $W$ signals the discrepancy with high confidence for $n \geq 10^3$, too.

**A step function.** The probability density function $p$ for our second example is a step function (a function which is constant on each interval in a particular partition of the real line into finitely many intervals). In particular, we define

$$p(x) = \begin{cases} 10^{-3}, & x \in (2k-1, 2k) \text{ for } k \in \{1, 2, \ldots, 999\} \\ 10^{-6}, & x \in (2k, 2k+1) \text{ for } k \in \{0, 1, 2, \ldots, 999\} \\ 0, & \text{otherwise} \end{cases} \quad \textbf{[13]}$$

for any $x \in \mathbb{R}$.

We compute the statistics $U$, $V$, and $W$ defined in Eqs. **3**, **4**, and **5** for two sets of i.i.d. draws, first for i.i.d. draws distributed according to $p$ defined in Eq. **13**, and then for i.i.d. draws from the uniform distribution on $(0, 1999)$. Table 3 displays numerical results.

For this example, the classical Kuiper statistic $U$ is unable to signal that the draws from the uniform distribution do not arise from $p$ defined in Eq. **13** for $n \leq 10^6$, at least not nearly as well as the modified Kuiper statistic $V$, which signals the discrepancy with high confidence for $n \geq 10^2$. The statistic $W$ does not signal the discrepancy for this example.

**A bimodal distribution.** The probability density function $p$ for our third example is

$$p(x) = \begin{cases} x/10100, & x \in [0, 100] \\ (101 - x)/101, & x \in [100, 101] \\ (x - 101)/101, & x \in [101, 102] \\ (202 - x)/10100, & x \in [102, 202] \\ 0, & \text{otherwise} \end{cases} \quad \textbf{[14]}$$

for any $x \in \mathbb{R}$. Figure 1 plots $p$.

We compute the statistics $U$, $V$, and $W$ defined in Eqs. **3**, **4**, and **5** for two sets of i.i.d. draws, first for i.i.d. draws distributed according to $p$ defined in Eq. **14**, and then for i.i.d. draws distributed according to the probability density function $q$ defined via the formula

$$q(x) = \begin{cases} x/101^2, & x \in [0, 101] \\ (202 - x)/101^2, & x \in [101, 202] \end{cases} \quad \textbf{[15]}$$

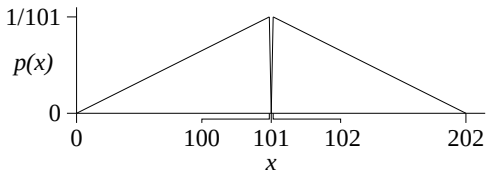


**Fig. 1.** The bimodal probability density function defined in Eq. **14**.

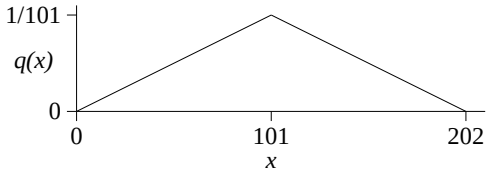


**Fig. 2.** The unimodal probability density function defined in Eq. **15**.

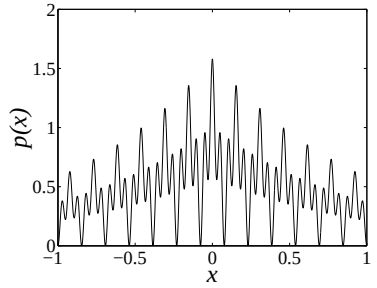


**Fig. 3.** The probability density function $p$ defined in Eq. **16**.

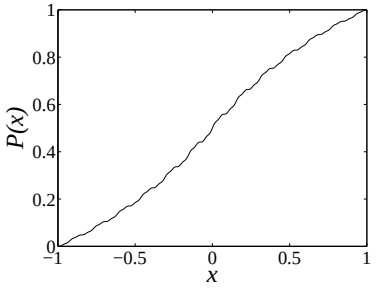


**Fig. 4.** The cumulative distribution function $P$ defined in Eq. **1** for $p$ in Eq. **16**.

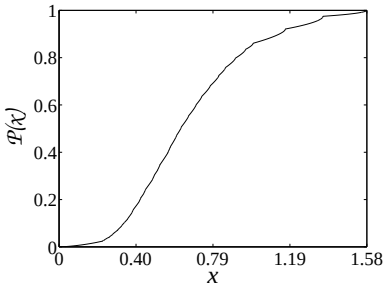


**Fig. 5.** The distribution function $\mathcal{P}$ defined in Eq. **2** for $p$ in Eq. **16**.

for any $x \in \mathbb{R}$. Figure 2 plots $q$. Table 4 displays numerical results.

For this example, the classical Kuiper statistic $U$ signals that the draws from $q$ defined in Eq. **15** do not arise from $p$ defined in Eq. **14** for $n \geq 10^5$, and the modified Kuiper statistic $V$ is inferior. The statistic $W$ signals the discrepancy with high confidence for $n \geq 10^4$.

**Table 2. A sawtooth wave**

| $n$ | $U_0$ | $U_1$ | $V_0$ | $V_1$ | $W_0$ | $W_1$ |
|---|---|---|---|---|---|---|
| $10^1$ | .13E1 | .12E1 | .11E1 | .14E1 | .24E1 | .49E–2 |
| $10^2$ | .12E1 | .18E1 | .10E1 | .21E1 | .37E0 | .45E–1 |
| $10^3$ | .82E0 | .79E0 | .13E1 | .81E1 ($10^{-54}$) | .18E1 | .10E–2 |
| $10^4$ | .12E1 | .17E1 | .13E1 | .25E2 ($10^{-7\mathrm{E}2}$) | .30E1 | .72E–4 |
| $10^5$ | .10E1 | .12E1 | .18E1 | .79E2 ($10^{-7\mathrm{E}3}$) | .18E0 | .34E–4 |
| $10^6$ | .81E0 | .14E1 | .12E1 | .25E3 ($10^{-7\mathrm{E}4}$) | .11E1 | .11E–4 |
| $10^7$ | .15E1 | .19E1 | .18E1 | .79E3 ($10^{-7\mathrm{E}5}$) | .13E1 | .38E–8 |

**Table 3. A step function**

| $n$ | $U_0$ | $U_1$ | $V_0$ | $V_1$ | $W_0$ | $W_1$ |
|---|---|---|---|---|---|---|
| $10^1$ | .11E1 | .12E1 | .32E–2 | .13E1 | .10E2 | .01E0 |
| $10^2$ | .11E1 | .18E1 | .10E–1 | .46E1 ($10^{-16}$) | .10E3 | .10E0 |
| $10^3$ | .10E1 | .81E0 | .32E–1 | .16E2 ($10^{-2\mathrm{E}2}$) | .10E1 | .10E1 |
| $10^4$ | .15E1 | .17E1 | .10E–1 | .50E2 ($10^{-3\mathrm{E}3}$) | .10E2 | .10E2 |
| $10^5$ | .11E1 | .12E1 | .22E–1 | .16E3 ($10^{-3\mathrm{E}4}$) | .10E3 | .10E3 |
| $10^6$ | .70E0 | .15E1 | .19E–1 | .50E3 ($10^{-3\mathrm{E}5}$) | .10E4 | .10E4 |
| $10^7$ | .65E0 | .33E1 ($10^{-8}$) | .12E–1 | .16E4 ($10^{-3\mathrm{E}6}$) | .10E5 | .10E5 |

**Table 4. A bimodal distribution**

| $n$ | $U_0$ | $U_1$ | $V_0$ | $V_1$ | $W_0$ | $W_1$ |
|---|---|---|---|---|---|---|
| $10^1$ | .11E1 | .14E1 | .11E1 | .14E1 | .11E0 | .98E–0 |
| $10^2$ | .15E1 | .15E1 | .11E1 | .12E1 | .37E0 | .19E–0 |
| $10^3$ | .11E1 | .10E1 | .10E1 | .13E1 | .21E1 | .70E–1 |
| $10^4$ | .12E1 | .19E1 | .15E1 | .11E1 | .70E0 | .68E–3 |
| $10^5$ | .10E1 | .33E1 ($10^{-8}$) | .11E1 | .18E1 | .88E0 | .40E–3 |
| $10^6$ | .65E0 | .99E1 ($10^{-82}$) | .68E0 | .57E1 ($10^{-25}$) | .14E0 | .25E–7 |
| $10^7$ | .89E0 | .31E2 ($10^{-1\mathrm{E}3}$) | .66E0 | .16E2 ($10^{-2\mathrm{E}2}$) | .29E0 | .25E–6 |

**Table 5. A differentiable density function**

| $n$ | $U_0$ | $U_1$ | $V_0$ | $V_1$ | $W_0$ | $W_1$ |
|---|---|---|---|---|---|---|
| $10^1$ | .12E1 | .74E0 | .11E1 | .11E1 | .14E1 | .11E–2 |
| $10^2$ | .14E1 | .11E1 | .18E1 | .30E1 ($10^{-5}$) | .13E1 | .17E–3 |
| $10^3$ | .15E1 | .14E1 | .92E0 | .57E1 ($10^{-26}$) | .51E0 | .22E–4 |
| $10^4$ | .86E0 | .22E1 ($10^{-3}$) | .12E1 | .16E2 ($10^{-2\mathrm{E}2}$) | .91E0 | .12E–5 |
| $10^5$ | .12E1 | .58E1 ($10^{-27}$) | .12E1 | .52E2 ($10^{-3\mathrm{E}3}$) | .72E0 | .12E–6 |

**A differentiable density function.** The probability density function $p$ for our fourth example is

$$p(x) = \begin{cases} C\, e^{-|x|}\,(2 + \cos(13\pi x) + \cos(39\pi x)), & x \in [-1, 1] \\ 0, & \text{otherwise} \end{cases} \qquad \textbf{[16]}$$

for any $x \in \mathbb{R}$, where $C \approx .4$ is the positive real number chosen such that $\int_{-\infty}^{\infty} p(x)\,dx = 1$. Figure 3 plots $p$. We evaluated numerically the corresponding cumulative distribution function $P$ defined in Eq. **1**, using the Chebfun package for Matlab described in [16]. Figure 4 plots $P$. We evaluated the distribution function $\mathcal{P}$ defined in Eq. **2** using the general-purpose scheme described in the appendix of [17] (which is also based on Chebfun). Figure 5 plots $\mathcal{P}$.

We compute the statistics $U$, $V$, and $W$ defined in Eqs. **3**, **4**, and **5** for two sets of i.i.d. draws, first for i.i.d. draws distributed according to $p$ defined in Eq. **16**, and then for i.i.d. draws distributed according to the probability density function $q$ defined via the formula

$$q(x) = \begin{cases} e^{-|x|}/(2 - 2\,e^{-1}), & x \in [-1, 1] \\ 0, & \text{otherwise} \end{cases} \qquad \textbf{[17]}$$

for any $x \in \mathbb{R}$. Table 5 displays numerical results.

For this example, the classical Kuiper statistic $U$ signals that the draws from $q$ defined in Eq. **17** do not arise from $p$ defined in Eq. **16** for $n \geq 10^4$, but not nearly as well as the modified Kuiper statistic $V$, which signals the discrepancy with high confidence for $n \geq 10^2$. The statistic $W$ signals the discrepancy with high confidence for $n \geq 10^2$, too.

**Remark 6.** *For all but the last example, the cumulative distribution function $P$ defined in Eq.* **1** *and the distribution function $\mathcal{P}$ defined in Eq.* **2** *are easy to calculate analytically; see, for example, [17]. However, as the last example illustrates, evaluating $P$ and $\mathcal{P}$ can in general require numerical algorithms such as the black-box schemes described in the appendix of [17].*

**Remark 7.** *For all numerical examples reported above, at least one of the modified Kuiper statistic $V$ or the "new" statistic $W$ is more powerful than the classical Kuiper statistic $U$, usually strikingly so. However, we recommend using all three statistics in conjunction, to be conservative. In fact, the statistics $V$ and $W$ of the present article are not able to discern certain characteristics of probability distributions that $U$ can, such as the symmetry of a Gaussian. The classical Kuiper statistic $U$ should be more powerful than its modification $V$ for any differentiable probability density function that has only one local maximum. For a differentiable probability density function that has only one local maximum, the "new" statistic $W$ amounts to an obvious test for outliers — nothing new (and far more subtle procedures for identifying outliers are available; see, for example, [18] and [19]). Still, as the above examples illustrate, $V$ and $W$ can be helpful with probability density functions that have multiple local maxima.*

## Conclusions and Generalizations

In this paper, we complemented the classical tests of the Kolmogorov-Smirnov type with tests based on the plain fact that it is unlikely to draw a random number whose probability is small, provided that the draw is taken from the same distribution used in calculating the probability (thus, if we draw a random number whose probability is small, then we can be confident that we did not draw the number from the same distribution used in calculating the probability). *Numerical Examples* above illustrates the substantial power of the supplementary tests, relative to the classical tests.

Needless to say, the method of the present paper generalizes straightforwardly to probability density functions of several variables. There are also generalizations to discrete distributions, whose cumulative distribution functions are discontinuous.

If the probability density function $p$ involved in the definition of the modified Kuiper test statistic $V$ in Eq. **4** takes on only finitely many values, then the confidence bounds of [3], [2], and Sections 14.3.3 and 14.3.4 of [1] are conservative, yielding lower than possible confidence levels that i.i.d. draws $X_1$, $X_2$, …, $X_{n-1}$, $X_n$ do not arise from $p$. It is probably feasible to compute the tightest possible confidence levels (maybe without resorting to the obvious Monte Carlo method), though we may want to replace $V$ with a better statistic when $p$ takes on only finitely many values; for example, when $p$ takes on only finitely many values, we can literally and explicitly rearrange $p$ to be nondecreasing on the shortest interval outside which it vanishes, and use the Kolmogorov-Smirnov approach on the rearranged $p$.

Even so, the confidence bounds of [3], [2], and Sections 14.3.3 and 14.3.4 of [1] for the modified Kuiper test statistic $V$ in Eq. **4** are sharp for many probability density functions $p$. For example, the bounds are sharp if, for every nonnegative real number $y$, the probability is 0 that $p(X) = y$, where $X$ is a random variable distributed according to $p$. This covers most cases of practical interest. In general, the tests of the present article are fully usable in their current forms, but may not yet be perfectly optimal for certain classes of probability distributions.

**ACKNOWLEDGMENTS.** We would like to thank Andrew Barron, Gérard Ben Arous, Peter Bickel, Sourav Chatterjee, Leslie Greengard, Peter W. Jones, Ann B. Lee, Vladimir Rokhlin, Jeffrey Simonoff, Amit Singer, Saul Teukolsky, Larry Wasserman, and Douglas A. Wolfe.

1. **Press W, Teukolsky S, Vetterling W, Flannery B (2007) *Numerical Recipes* (Cambridge University Press, Cambridge, UK), 3rd ed.**
2. **Stephens MA (1970) Use of the Kolmogorov-Smirnov, Cramer–Von-Mises and related statistics without extensive tables. *J Roy Statist Soc Ser B* 32:115–122.**
3. **Stephens MA (1965) The goodness-of-fit statistic $V_n$: Distribution and significance points. *Biometrika* 52:309–321.**
4. **Barron AR (1989) Uniformly powerful goodness of fit tests. *Ann Statist* 17:107–124.**
5. **Bickel PJ, Rosenblatt M (1973) On some global measures of the deviations of density function estimates. *Ann Statist* 1:1071–1095.**
6. **Fan J (1996) Test of significance based on wavelet thresholding and Neyman's truncation. *J Amer Statist Soc* 91:674–688.**
7. **Hollander M, Wolfe DA (1999) *Nonparametric Statistical Methods* (Wiley, New York), 2nd ed.**
8. **Inglot T, Ledwina T (1996) Asymptotic optimality of data-driven Neyman's tests for uniformity. *Ann Statist* 24:1982–2019.**
9. **Khamis HJ (2000) The two-stage $\delta$-corrected Kolmogorov-Smirnov test. *J Appl Statist* 27:439–450.**
10. **Rayner JCW, Thas O, Best DJ (2009) *Smooth Tests of Goodness of Fit* (Wiley), 2nd ed.**
11. **Reschenhofer E (2008) Combining generalized Kolmogorov-Smirnov tests. *InterStat* June:1–15.**
12. **Simonoff JS (1996) *Smoothing Methods in Statistics* (Springer-Verlag, New York).**
13. **Wasserman L (2003) *All of Statistics* (Springer).**
14. **Bickel PJ, Ritov Y, Stoker TM (2006) Tailor-made tests for goodness of fit to semiparametric hypotheses. *Ann Statist* 34:721–741.**
15. **Stein EM, Weiss G (1971) *Introduction to Fourier Analysis on Euclidean Spaces* (Princeton University Press, Princeton, NJ).**
16. **Trefethen LN, Hale N, Platte RB, Driscoll TA, Pachón R (2009) Chebfun, version 3 (http://www.maths.ox.ac.uk/chebfun, Oxford University).**
17. **Tygert M (2010) Statistical tests for whether a given set of independent, identically distributed draws does not come from a specified probability density (http://arxiv.org, arXiv), Tech Rep 1001.2286.**
18. **Simonoff JS (1987) Outlier detection and robust estimation of scale. *J Stat Comput Simul* 27:79–92.**
19. **Davies L, Gather U (1993) The identification of multiple outliers. *J Amer Statist Assoc* 88:782–792.**