Abstract

This paper takes a look at omnibus tests of goodness-of-fit in the context of reweighted Anderson-Darling tests and makes three fold contributions. The first contribution is to provide a geometric understanding. It is argued that the test statistic with minimum variance can serve as a good general-purpose test. The second contribution is to propose better omnibus tests, called circularly symmetric tests and obtained by circularizing reweighted Anderson-Darling tests. A limited but arguably convincing simulation study on finite-sample performance demonstrates that the circularized tests outperform their parent methods. The third contribution is to establish new large-sample results. It is shown that like Anderson-Darling, the minimum-variance test statistic and two circularly symmetric test statistics under the null has the same distribution as that of a weighted sum of an infinite number of independent squared normal random variables.

Key Words: Circulant matrices; Cramér-von Mises; Gaussian processes; Reweighted Anderson-Darling; Sturm-Liouville equation.

1 Introduction

The problem of determining whether a sample of \( n \) observations \( X_1, \ldots, X_n \) can be considered as a sample from a given continuous distribution \( F(x) \), known as goodness-of-fit, is theoretically fundamental. It is also practically important, especially for contemporary big-data analysis, for model building and checking in particular and non-parametric inference in general. The methodology development for assessing goodness-of-fit has been a good part of statistical research in the past century. It can be traced back to Pearson’s chi-square test (Pearson, 1900) and has made available many influential methods, including Kolmogorov-Smirnov test (Kolmogorov 1933; Smirnov 1939), Cramér-von Mises criterion (Cramér 1928; von Mises 1928), Anderson-Darling test (Anderson and Darling 1952, 1954), Shapiro-Wilk test (Shapiro and Wilk 1965), and Zhang test (Zhang 2002); See (Lehmann and Romano 2005 p. 629-630) for a comprehensive list of references. Among these classical tests, Anderson-Darling and Zhang have been perceived as powerful (see, e.g., Sinclair and Spurr 1988; Zhang 2010).

Anderson-Darling test (Anderson and Darling 1952) is defined as a weighted empirical distribution statistic

\[
A_n^2(w) = n \int_{-\infty}^{\infty} [F_n(x) - F(x)]^2 w(x) dx
\]

with the null distribution \( F(.) \) and the weight function \( w(x) = \frac{1}{f(x)(1-F(x))} \), where \( F_n(.) \) denotes
the usual empirical distribution function:

\[ F_n(x) = \frac{k}{n} \quad \text{if } k \text{ observations are } \leq x. \]

In comparing Anderson-Darling and Cramér-von Mises, Anderson and Darling (1952) wrote:

A statistician may prefer to use this weight function \( \psi(x) = \frac{1}{[F(x)(1 - F(x))] } \) when he feels that \( \psi(x) = 1 \) does not give enough weight to the tails of the distribution \( F(x) \).

While this is true, the comparison is nevertheless only relative. The most important to an applied statistician is perhaps that the question ‘what would be a default or all-purpose test that could be considered relatively neutral regarding the location of deviations from the hypothesized distribution?’ remains, however, to be answered. This paper takes a look at this so-called omnibus testing problem and aims to three goals.

The first goal of this paper is to develop geometric intuitions and corresponding mathematical theory toward an answer to the above question by considering a class of reweighted Anderson-Darling tests. It establishes an explicit one-to-one correspondence between the weights and their focal directions of distributional deviations of \( F^*(.) \) from \( F(.) \). It is found that the weights that produce the test statistic with minimal variance put equal focuses on all the standardized deviations. As a result, we take the corresponding test as a general-purpose test. This arguably optimal weights-based test is found to be practically equivalent to the Zhang test, which has been commonly perceived powerful. It should noted that while the existing simulation-based empirical results are helpful, our conclusions about the performance is based on our geometric arguments and the corresponding theoretical results.

The second goal is to explore better omnibus tests. Is is recognized that existing powerful methods suffer from the confounded effect of locations and frequencies in the deviations from the null hypothesis. This motivates the method of circularization to create circularly symmetric tests by circularizing reweighted Anderson-Darling tests. Two types of circularization are considered: one is obtained by taking the average of the corresponding scan statistics and the other by using the maximum. A simple but arguably convincing simulation study in Section 5 on finite-sample performance demonstrates that the circularized Zhang method outperforms the circularized Anderson-Darling and that the circularized tests outperform their parent methods.

The final goal is to establish new large-sample results. It is found that like Anderson-Darling, the test statistics under the null have the same distribution as that of a weighted sum of an infinite number of independent squared normal random variables. These theoretical results are shown numerically to be useful for large sample-based approximations.

The rest of the paper is arranged as follows. Section 2 introduces basic notations and the class of reweighted Anderson-Darling test statistics. Section 3 develops statistical intuitions for understanding reweighted Anderson-Darling tests. The default or optimal weights are then defined accordingly, followed by an investigation on finite-sample cases and a large sample theory-based solution. Section 4 discusses the difficulties suffered by existing powerful methods and proposes circularly symmetric tests. Section 5 considers a simple simulation study, which demonstrates that the circularized Zhang method outperforms the circularized Anderson-Darling and that the circularized tests outperform their parent methods. Section 6 discusses the limiting distributions of three proposed test statistics. Section 7 concludes with a few remarks.
2 Reweighted Anderson-Darling Tests

2.1 Reweighted Anderson-Darling tests

The basic setting for the theoretical investigation is that the independent and identically distributed random variables $X_1, ..., X_n$ have a specified continuous distribution $F(x)$, $x \in \mathbb{R}$. Denote by $X_{(1)} \leq \ldots \leq X_{(n)}$ the corresponding order statistics. This set of order statistics or the corresponding order statistics $U_{(1)} = F(X_{(1)}) \leq \ldots \leq U_{(n)} = F(X_{(n)})$ are sufficient for inference about $F(.)$, especially when inference about unknown $F(.)$ is of interest. It is well-known and easy-to-prove that the sampling distribution of $U_{(1)}, ..., U_{(n)}$ is that of a sorted uniform sample of size of $n$. In the context of hypothesis testing, we write the null hypothesis as

$$H_0 : F^*(x) = F(x) \quad \text{for all } x \in \mathbb{R}$$

and the alternative as

$$H_1 : F^*(x) \neq F(x) \quad \text{for some } x \in \mathbb{R}$$

where $F^*(.)$ stands for the true distribution of the observed sample $X_1, ..., X_n$. So it should be noted that when relevant in what follows, the distribution of the sorted $U_{(i)}$ is that under the null hypothesis $H_0$.

Let $a_i = \frac{i - \frac{3}{8}}{n}$ for $i = 1, ..., n$. The Anderson-Darling statistic can be written simply and equivalently as

$$\bar{A}_n^2 = -2 \sum_{i=1}^{n} \left[ a_i \ln \frac{U_{(i)}}{\hat{a}_i} + (1 - a_i) \ln \frac{1 - U_{(i)}}{1 - a_i} \right], \quad (2.1)$$

making it statistically more intuitive in terms of the finite number of sufficient statistics $U_{(1)}, ..., U_{(n)}$. In this paper, for theoretical convenience for $\mu_i = E(U_{(i)}) = \frac{i}{n+1}$, we replace $\hat{a}_i$ in (2.1) with $\mu_i$ and consider the slightly modified version:

$$W_n^2 = -2 \sum_{i=1}^{n} \left[ \mu_i \ln \frac{U_{(i)}}{\mu_i} + (1 - \mu_i) \ln \frac{1 - U_{(i)}}{1 - \mu_i} \right]. \quad (2.2)$$

That is, this is done in an analogy with methods using the alternative definition of empirical distribution

$$F_n(x) = \frac{|\{i : X_i \leq x\}|}{n + 1}, \quad (x \in (-\infty, \infty)),$$

where $|\{i : X_i \leq x\}|$ is the number of $X_i$s that are less than or equal to $x$.

The modified version (2.2) has a very simple intuitive interpretation. Note that the marginal probability density function (pdf) of $U_{(i)}$ is Beta($i$, $n+1-i$), the Beta distribution with the two shape parameters $i$ and $n+1-i$ (see, e.g., David and Nagaraja, 2004, p.14). Let $Z_i = \ln \frac{U_{(i)}}{1-U_{(i)}}$, the logit transformation of $U_{(i)}$. The $i$-th summand of $W_n^2$ is proportional to the negative log probability density function (pdf). This implies that Anderson-Darling test statistic is approximately the average of squares of standardized $Z_i$s, which is stochastically small under the null hypothesis and large under the alternative. This can be seen more easily with the following approximation to the $i$-th summand of $W_n^2$ via the Taylor expansion in terms of $U_{(i)}$ at $U_{(i)} = \mu_i$:

$$Y_i \equiv -2 \left[ \mu_i \ln \frac{U_{(i)}}{\mu_i} + (1 - \mu_i) \ln \frac{1 - U_{(i)}}{1 - \mu_i} \right] \approx \frac{1}{(n+2)} \frac{(U_{(i)} - \mu_i)^2}{\text{Var}(U_{(i)})} \quad (2.3)$$
where $\text{Var}(U_{(i)}) = \frac{\mu_i(1-\mu_i)}{n+2}$, under the null and going to zero as $n \to \infty$. Recall that Cramér-von Mises test statistic is basically the straight average of the squared deviations $(U_{(i)} - \mu_i)^2$, 

$$
\frac{1}{12n} + \sum_{i=1}^{n} (U_{(i)} - \tilde{a}_i)^2 \approx \frac{1}{12n} + \sum_{i=1}^{n} (U_{(i)} - \mu_i)^2.
$$

(2.4)

So compared to Cramér-von Mises, Anderson-Darling is the weighted average of the squared deviations $(U_{(i)} - \mu_i)^2$ with the weights inversely proportional to the variance of the deviations $U_{(i)} - \mu_i$.

Notice that the deviations $U_{(i)} - \mu_i$ and, thereby, their squared versions $(U_{(i)} - \mu_i)^2$ near the central area of $F(x)$ are more correlated than those in the tails. It is worth considering to weight the tail areas even more than Anderson-Darling. This motivates us to consider the following class of reweighted Anderson-Darling test statistics:

$$
R_n^2(w) = -2 \sum_{i=1}^{n} w_i \left[ \mu_i \ln \frac{U_{(i)}}{\mu_i} + (1 - \mu_i) \ln \frac{1 - U_{(i)}}{1 - \mu_i} \right]
$$

(2.5)

where $w_i \geq 0$ for $i = 1, ..., n$. The special equal-weight case of $w_i = 1$ corresponds to the slightly modified Anderson-Darling test statistic (2.2).

The default or optimal weights, defined for terminology convenience and studied in detail in Section 3, are found to be

$$
w_i \propto \frac{1}{\mu_i(1-\mu_i)} \quad (i = 1, ..., n)
$$

(2.6)

for large $n$ and weight tails slightly more for small $n$. This leads to the following test

$$
R_n^2 = -2C_n \sum_{i=1}^{n} \frac{1}{\mu_i(1-\mu_i)} \left[ \mu_i \ln \frac{U_{(i)}}{\mu_i} + (1 - \mu_i) \ln \frac{1 - U_{(i)}}{1 - \mu_i} \right]
$$

(2.7)

where the rescaling constant $C_n = \frac{1}{2 \sum_{i=1}^{n} \frac{1}{\mu_i}}$ is taken for the preference of $E(R_n^2) \approx 1$ (see Section 6.1).

Clearly, this test is practically equivalent to the test of Zhang (2002):

$$
Z_A = -\sum_{i=1}^{n} \left[ \frac{\ln(U_{(i)})}{n-i+\frac{1}{2}} + \frac{\ln(1-U_{(i)})}{i-\frac{1}{2}} \right],
$$

(2.8)

which is obtained via weighting likelihood ratios for individual $F(t)$ instead of $[F_n(x) - F(x)]^2$ in (1.1) with the choice of the weight function $1/[F(x)(1 - F(x))]$. It should be noted that the Zhang test has been commonly perceived as powerful (see, e.g., Carmen Pardo et al., 2022, Zhang, 2002, 2010). The good performance of $R_n^2$ relative to $W_n^2$ is also demonstrated from a different perspective in Section 3. This helps to see the performance of $R_n^2$, in addition to the geometric interpretation provided in Section 3. Likewise, our new asymptotic results established in Section 6.1 can be applied to the Zhang test $Z_A$. In addition to these theoretical contributions, the geometric intuitions and the corresponding theoretical investigation in the next section shed light on why $R_n$ is chosen to be a default test and when it has the best performance.

3 Optimal Weights for Minimum Variance

3.1 Intuitions and definition

The discussion so far has been mainly focused on understanding of $U_{(i)}$ as a pivotal quantity, that is, its distribution under the null hypothesis. To help see what evidence against the null hypothesis
the $Y_i$s capture, denote by $x(i)$ the $\mu_i$-th quantile of $F^*(x)$, that is, $F^*(x(i)) = \mu_i$. Then

$$E(Y_i) - 1 \approx \frac{[F(x(i)) - F^*(x(i))]^2}{\mu_i(1 - \mu_i)}, \quad (3.1)$$

a type of signal-to-noise ratio. This implies that the weighted Anderson-Darling $R^2_n(w)$ deals with the $n$ pieces of unknown quantities $[F(x(i)) - F^*(x(i))]^2 / [\mu_i(1 - \mu_i)]$ by capturing the weighted sum of signal-to-noise ratios:

$$E[R^2_n(w)] - \sum_{i=1}^{n} w_i \approx \sum_{i=1}^{n} w_i \frac{[F(x(i)) - F^*(x(i))]^2}{\mu_i(1 - \mu_i)}. \quad (3.2)$$

For a geometric interpretation, consider the signal-to-noise ratio

$$\delta_i = \frac{[F(x(i)) - F^*(x(i))]^2}{\mu_i(1 - \mu_i)} \quad (3.3)$$

for $i = 1, \ldots, n$. Suppose that we use a prespecified $n$-vector $\delta = (\delta_1, \ldots, \delta_n)'$ to represent an alternative of interest in the context of the hypothesis. Then we can consider to use the weights that maximize the signal-to-noise ratio for the reweighted Anderson-Darling. This solution exists and is summarized into the following lemma.

**Lemma 1.** Let $\delta = (\delta_1, \ldots, \delta_n)'$, where $\delta_i$ is defined in (3.3). Denote by $\Sigma$ the covariance matrix of $Y = (Y_1, \ldots, Y_n)'$, where $Y_i$ is defined in (2.3). If $\delta \neq 0$, then for all $w \neq 0$

$$\frac{(w'\delta)^2}{w'\Sigma w} \leq \delta\Sigma^{-1}\delta$$

with the equality holds if and only if $w \propto \Sigma^{-1}\delta$.

Figure 1: Focal directions of Anderson-Darling, defined as $\Sigma 1$ in Subsection 3.1 for $n = 10$, 50, and 100.

This is a familiar mathematical result and can be proved straightforwardly by applying Cauchy-Schwartz inequality theorem to the two vectors $\Sigma^2 w$ and $\Sigma^{-1}\delta$. The covariance matrix of $Y = (Y_1, \ldots, Y_n)'$ is given in Theorem 1 below in Subsection 3.2. This allows for a geometric interpretation regarding the performance of the test statistic $R^2_n(w)$: for any given weights $w$, the test statistic is
the most powerful when the direction $\delta$ is proportional to $\Sigma w$. We call $\Sigma w$ the focal direction of the weighted test $R^2_n(w)$. For example, the focal direction of Anderson-Darling is $\delta \propto \Sigma 1$, which is shown in Figure [1] for the three cases of $n = 10, 50$, and $100$. Clearly, due to the strong correlations among $U(i)$’s, Anderson-Darling effectively focuses on the central area more than the tail areas.

It is seen thus far that the challenge of goodness-of-fit is due to the fact that it essentially involves simultaneous inference on multiple parameters $F^*(X(i))$, $i = 1, \ldots, n$. So there are no uniformly optimal weights in general. Consequently, we can only consider optimal weights by considering a certain type of performance, such as typical or average performance. Consider the uniformly optimal weights in general. Consequently, we can only consider optimal weights by

\[ F \text{ involves simultaneous inference on multiple parameters } F^*(X(i)), \ i = 1, \ldots, n. \]

So there are no uniformly optimal weights in general. Consequently, we can only consider optimal weights by considering a certain type of performance, such as typical or average performance. Consider the case where $\delta_i$ are exchangeable and summarized as $\delta \propto 1$, interpreted in practice as a type of average from experiment to experiment. It follows from Lemma [1] that the optimal weights $w$ in this case are the optimal weights that minimize the variance of $R^2_n(w)$. Formally, for terminology convenience, we define optimal weights for minimum variance as follows.

**Definition 3.1.** The weights $w^{(\text{optim})}$ satisfying $\sum_{i=1}^n w_i = c$ for some positive $c$ are called optimal for minimum variance if

\[ w^{(\text{optim})} = \arg \min_{\sum_{i=1}^n w_i = 1} \text{Var} \left( R^2_n(w) \right). \]

**Remark 3.1.** The above argument leading to Definition [3.1] could also be explained loosely as follows. Suppose that from experiment to experiment, $\delta$ appears like a random variable with the mean vector $\tau 1$ and covariance matrix $\epsilon^{-1} \Sigma$ for some $\tau > 0$ and $\epsilon > 0$, written as $\delta \sim (\tau 1, \epsilon^{-1} \Sigma)$, and that similarly $Y \sim (\delta, \Sigma)$. Then $w^T Y \sim (\tau w^T 1, [\epsilon^{-1} + 1] w^T \Sigma w)$. From the argument above and Lemma [4] we see that the particular weights are defined in Definition [3.1] maximize $(\tau w^T 1)^2 / ([\epsilon^{-1} + 1] w^T \Sigma w)$.

We discuss the optimal weights below in Subsections [3.2] and [3.3] for finite sample and large sample cases respectively.

### 3.2 Finite sample cases

The optimal weights $w^{\text{opt}}$ depend on the evaluation of the variance of $R^2_n(w)$. Since $R^2_n(w)$ is linear in $\ln (U(i))$’s and $\ln (1 - U(i))$’s, the variance of $R^2_n(w)$ can be obtained exactly by making use of the following results.

**Theorem 1.** Let $U(1) < \ldots < U(n)$ be the sorted uniforms of size $n$. Denote by $\psi(x)$ and $\psi_1(x)$ the digamma and trigamma functions respectively. Then

(a) $E[\ln (U(i))] = \psi(i) - \psi(n + 1), \ E[\ln (1 - U(i))] = \psi(n + 1 - i) - \psi(n + 1), \ \text{Var}[\ln (U(i))] = \psi_1(i) - \psi_1(n + 1)$, and $\text{Var}[\ln (1 - U(i))] = \psi_1(n + 1 - i) - \psi_1(n + 1)$ for $i = 1, \ldots, n$;

(b) $\text{Cov}[\ln (U(i)), \ln (U(j))] = \psi_1(j) - \psi_1(n + 1)$ and $\text{Cov}[\ln (1 - U(i)), \ln (1 - U(j))] = \psi_1(n + 1 - i) - \psi_1(n + 1)$ for all $1 \leq i < j \leq n$; and

(c) $\text{Cov}[\ln (U(i)), \ln (1 - U(j))] = -\psi_1(n + 1)$ and

\[
\text{Cov}[\ln(1 - U(i)), \ln(U(j))] = \frac{\Gamma(n + 1)}{\Gamma(i)} \sum_{k=1}^{\infty} \frac{\Gamma(i + k)}{k \Gamma(n + 1 + k)} [\psi(n + 1 + k) - \psi(j + k)] - [\psi(n + 1 - i) - \psi(n + 1)] [\psi(j) - \psi(n + 1)]
\]

for all $1 \leq i < j \leq n$. 

6
Figure 2: Finite-sample exact results on optimal weights for three cases $n = 10, 50,$ and $100$. The slope is regression coefficient of the least-squares fit of the optimal weights on the variance of $U(i)$, both in log scale.

The proof of Theorem 1 is provided in Appendix A.1. The optimal weights for $n = 10, 50,$ and $100$ are shown in Figure 2. These numerical results clearly suggest that the optimal weights are almost inversely proportional to the variances of $U(i)$, that is,

$$w_i \propto \frac{1}{\mu_i(1 - \mu_i)} = \frac{(n + 1)^2}{i(n + 1 - i)}$$

for $i = 1, \ldots, n$. These optimal results are in fact asymptotically exact, which is discussed in the next subsection. The numerical results in Figure 2 show that for small $n$, the optimal weights weight slightly more on the tails of $F(x)$, but they appear to converge very quickly to (3.4).

### 3.3 Large sample results

Since $\text{Var}(U(i)) = \frac{1}{n+2}\mu_i(1 - \mu_i)$ converges to zero as $n \to \infty$, we use the delta method to approximate $R_n^2(w)$ in terms of $U(i)$’s. This allows us to work conveniently with $U(i)$’s for investigating the large-sample behavior of $R_n^2$. Since the coefficient of the corresponding Taylor expansion for the first-order $U(i) - \mu_i$ is zero, we have

$$R_n^2(w) \approx \frac{1}{n+2} \sum_{i=1}^{n} \frac{w_i}{\text{Var}(U(i))} (U(i) - \mu_i)^2 = \sum_{i=1}^{n} \frac{w_i}{\mu_i(1 - \mu_i)} (U(i) - \mu_i)^2.$$ 

It is seen that in this case, to find the optimal weights we need to work with the covariance matrix of the vector of squared $(U(i) - \mu_i)$’s. The results on this covariance matrix are summarized into the following theorem, with the proof given in Appendix A.2.

**Theorem 2.** Let $U(1) < \ldots < U(n)$ be the sorted uniforms of size $n$. Then

$$\text{Cov}[(U(i) - \mu_i)^2, (U(j) - \mu_j)^2] = \frac{2\mu_i^2(1 - \mu_j)^2}{(n+2)(n+3)} + \frac{\mu_i(1 - \mu_j)}{(n+2)(n+3)} \left( \frac{3(1 - 3\mu_i)(2 - 3\mu_j)}{n+4} - \frac{1 - \mu_i}{(n+2)} \mu_j \right)$$

hold for all $1 \leq i \leq j \leq n.$
It is not hard to see that as $n \to \infty$, the scaled covariance $n^2 \text{Cov} \left[ (U_i' - \mu_i)^2, (U_j' - \mu_j)^2 \right]$ converges to $2\mu_i^2(1 - \mu_j)^2$. This is closely related to the mathematical treatment of Anderson and Darling (1952) using the limiting process of $B(i/(n+1)) = \sqrt{n + 2}[U(i) - E(U(i))], i = 1, ..., n$, which is a Gaussian process or, more exactly, the Brownian bridge. In this case, we can easily obtain the corresponding results for the covariance structure of the limiting process, which is characterized in the following theorem.

**Theorem 3.** The limiting process of $B(i/(n+1)) = \sqrt{n + 2}[U(i) - E(U(i))]$ is the Brownian bridge. For the Brownian bridge,

$$\text{Cov} [B(s), B(t)] = s(1 - t)$$

and

$$\text{Cov} [B^2(s), B^2(t)] = 2s^2(1 - t)^2$$

holds for all $0 \leq s \leq t \leq 1$.

The proof is given in Appendix A.3. From this result, we can obtain the asymptotic optimal weights or, more exactly, the optimal weight function. The result is summarized into the following theorem, with the proof given in Appendix A.4.

**Theorem 4.** In the limit as $n \to \infty$, the optimal weight function is given by

$$w(t) \propto \frac{1}{t(1 - t)} \quad (t \in [\varepsilon, 1 - \varepsilon]),$$

(3.5)

for all $\varepsilon \in \left(0, \frac{1}{2}\right]$.

This result provides the theoretical support to the use of the weights defined in (2.6) as the optimal weights in practice for all sample size $n$.

### 4 Circularly Symmetric Tests

#### 4.1 Intuition and definition

Since the distributional deviations in the central areas of $F(x)$ captured by the $U(i)$’s are averaged together with those in the tail areas, efficiency of these methods is questionable in such cases. This implies that the above methods are all based on sorted uniforms and, thereby, suffer from the confounded effect of different locations and various signal frequencies in the deviations from the null hypothesis. This motivates the idea of circularization to eliminate the location effect so that we allow the weights to focus on the various signal frequencies.

The technique of circularization is straightforward and is described as follows. Let $U(0) = 0$ and let $U(n + 1) = 1$. Extend the uniform spacings $D_i = U(i) - U(i-1)$ as a circular process at $n + 1$ locations equally spaced on a circle or on the index set of $\{0, ..., 2(n + 1)\}$. That is,

$$D_{n+1+i} = U((n+1)+i) - U((n+1)+(i-1)) = U(i) - U(i-1) = D_i$$

for $i = 1, ..., n + 1$. Define the circular counterparts of $U(i) = U(i) - U(0)$:

$$U_{(c)}(i) = U_{(c+i)} - U_{(c)} \quad (c = 0, ..., n; i = 1, ..., n + 1).$$

(4.1)
Now we define the circularly symmetric version of \((2.5)\) as the average of the \(n + 1\) location-shifted \(B(U^{(c)}, w)\) test statistics:

\[
C(U; w) = -\frac{2}{n + 1} \sum_{c=0}^{n} \sum_{i=1}^{n} w_i \left[ \mu_i \ln \frac{U^{(c)}(i)}{\mu_i} + (1 - \mu_i) \ln \frac{1 - U^{(c)}(i)}{1 - \mu_i} \right].
\] (4.2)

Here in this paper we focus on the circularized versions of \(W_n^2\) and \(R_n^2\):

\[
\tilde{W}_{n}^2 \equiv -\frac{2}{n + 1} \sum_{c=0}^{n} \sum_{i=1}^{n} \left[ \mu_i \ln \frac{U^{(c)}(i)}{\mu_i} + (1 - \mu_i) \ln \frac{1 - U^{(c)}(i)}{1 - \mu_i} \right]
\] (4.3)

and

\[
\tilde{R}_n^2 \equiv -\frac{1}{(n + 1)} \sum_{k=1}^{n} \sum_{c=0}^{n} \sum_{i=1}^{n} \frac{1}{\mu_i(1 - \mu_i)} \left[ \mu_i \ln \frac{U^{(c)}(i)}{\mu_i} + (1 - \mu_i) \ln \frac{1 - U^{(c)}(i)}{1 - \mu_i} \right].
\] (4.4)

The normalizing constants in (4.3) and (4.4) are taken in such a way that the expectations of these test statistics are about one for \(n\) large; See Remarks 6.2 and 6.3 in Section 6.2.

Alternative ways of circularization can be also worth consideration. For example, one may use the maximum of the scan statistics \(B(U^{(c)}, w)\)’s instead of the average, that is, \(\hat{C}(U, w) = \max_c B(U^{(c)}, w)\), using the notations in Section 4. Accordingly, we introduce the corresponding circularized versions of \(\dot{W}_n^2\) and \(\dot{R}_n^2\):

\[
\dot{W}_{n}^2 \equiv -\frac{2}{n + 1} \max_{c \in \{0,\ldots,n\}} \sum_{i=1}^{n} \left[ \mu_i \ln \frac{U^{(c)}(i)}{\mu_i} + (1 - \mu_i) \ln \frac{1 - U^{(c)}(i)}{1 - \mu_i} \right]
\] (4.5)

and

\[
\dot{R}_n^2 \equiv -\frac{1}{\sum_{k=1}^{n} \max_{c \in \{0,\ldots,n\}} \sum_{i=1}^{n} \frac{1}{\mu_i(1 - \mu_i)} \left[ \mu_i \ln \frac{U^{(c)}(i)}{\mu_i} + (1 - \mu_i) \ln \frac{1 - U^{(c)}(i)}{1 - \mu_i} \right].
\] (4.6)

A simple simulation study is done in next section to compare the performance of \(\dot{W}_n^2\) and \(\dot{R}_n^2\) and their circularized versions. Large sample results for understanding and numerical approximation are given in Section 6.2 for \(\tilde{W}_{n}^2\) and \(\tilde{R}_n^2\). Large sample results for \(\dot{W}_n^2\) and \(\dot{R}_n^2\) appear challenging and are expected to be considered elsewhere.

### 5 Power Comparison: a Simulation Study

We focus on our investigation on the performance of \(W_n^2; R_n^2\), and their circularized versions, \(\tilde{W}_n^2\) and \(\tilde{R}_n^2\), for a class of situations where the deviations of \(F^*(\cdot)\) from \(F(\cdot)\) are locally smooth but in different locations. For this, we use, without loss of generality, the standard uniform \(\text{Unif}(0,1)\) as \(F(\cdot)\) and consider the class of \(F^*(\cdot)\)’s obtained by simple local perturbations. More precisely, \(F^*(\cdot)\) is given by the following probability density function (pdf)

\[
f_{\eta, \sigma, \tau}(x) = 1 + \tau 1_{(\eta-\sigma, \eta]}(x) - \tau 1_{(\eta, \eta+\sigma)}(x) \quad (0 < x < 1)
\] (5.1)
where $\sigma > 0$, $0 \leq \eta - \sigma, \eta + \sigma \leq 1$, $0 \leq \tau \leq 1$, and $1_A(x)$ is the indicator function of the subset $A$, that is, $1_A(x) = 1$ if $x \in A$ and $1_A(x) = 0$ otherwise. The corresponding $F^*(\cdot)$ has the pdf

$$F_{\eta,\sigma,\tau}(x) = \begin{cases} x, & \text{if} \ p \in (0, \eta - \sigma]; \\ x + \tau(x - \eta + \sigma), & \text{if} \ x \in (\eta - \sigma, \eta]; \\ x - \tau(x - \eta - \sigma), & \text{if} \ x \in (\eta, \eta + \sigma); \\ x, & \text{if} \ x \in [\eta + 1), \end{cases} \quad (0 < x < 1) \quad (5.2)$$

with the inverse cdf

$$F^{-1}_{\eta,\sigma,\tau}(p) = \begin{cases} p, & \text{if} \ p \in (0, \eta - \sigma]; \\ p - \tau(p - \eta + \sigma), & \text{if} \ p \in (\eta - \sigma, \eta + \tau\sigma]; \\ p + \frac{\tau(p - \eta - \sigma)}{1 - \tau}, & \text{if} \ p \in (\eta + \tau\sigma, \eta + \sigma]; \\ p, & \text{if} \ p \in [\eta + 1), \end{cases} \quad (0 < p < 1). \quad (5.3)$$

Figure 3: Power comparisons in Section 5. The title $\{\tau, (\eta - \sigma, \eta + \sigma)\}$ of each plots represents the magnitude $\tau$ and location $(\eta - \sigma, \eta + \sigma)$ where $F^*(\cdot)$ deviates from $F(\cdot)$. The performance in terms of CDF of Fisher’s P-value are shown by solid curves for $R_n^2$ and dotted for $W_n^2$ in the four plots in the upper panel. The four plots in the lower penal show the corresponding results for $\tilde{R}_n^2$ and $\tilde{W}_n^2$. The diagonal line is the uniform reference, corresponding to tests that have no power.

For a simple but representative numerical comparison of $W_n^2$, $R_n^2$ and their circularized versions $\tilde{W}_n^2$ and $\tilde{R}_n^2$, we consider two scenarios. In the first scenario, we have a small interval of length 0.1 with an appropriate magnitude $\tau = 0.99$ large enough for visible differences of the performance. The tail and central locations are specified by $\eta = 0.05$ and 0.5. The performance of $W_n^2$ (dotted curve) and $R_n^2$ (solid curve) in terms of the cumulative distribution function (CDF) of their Fisher’s
P-value is shown by the first two plots on the upper panel of Figure 3. The performance of their circularized versions is shown by the first two plots on the lower panel of Figure 3. We see a dramatical improved balance of performance with respect to the unknown locations of deviations. Clearly, the performance of the circularized versions are location invariant, as expected.

In the second scenario, we have a large interval of length 0.5 with an appropriate magnitude $\tau = 0.25$. One may reach to the same conclusion as that in the first scenario. In summary, the results in Figure 3 shows clearly that the circularized Zhang method outperforms the circularized Anderson-Darling and that the circularized tests outperform their parent methods.

The corresponding comparison is made and shown in Figure 4 on the performance of $\check{W}_n^2$ and $\check{R}_n^2$, defined in (4.5) and (4.6), respectively. The results seem to suggest that $\check{W}_n^2$ and $\check{R}_n^2$ improve the performance when signals are of high frequency, while they are comparable to $\check{W}_n^2$ and $\check{R}_n^2$ for signals of low frequency.

6 Large-Sample Results

6.1 The Limiting Distribution of $R_n^2$

In this section, we investigate the asymptotic distribution of $R_n^2$ by taking the approach of Anderson and Darling (1952) and the extended result of their Theorem 4.1. In the present case, the kernel function is

$$\sqrt{w(s)} \sqrt{w(t)}[\min(s, t) - st] = \frac{1}{s(1 - s)} \frac{1}{t(1 - t)}[\min(s, t) - st] \quad (s, t \in [\varepsilon, 1 - \varepsilon])$$
where \( \varepsilon \) is a small positive number, say \( \varepsilon = 1/[2(n + 1)] \) in the context of a given sample. We use a small \( \varepsilon \) to rule out index values near the two end points of the interval \((0, 1)\) because the kernel function is unintegrable. This does not mean we cannot consider the limiting distribution of \( R^2_n \) for understanding the large-sample behavior of \( R^2_n \) and for large-sample approximation to the distribution of \( R^2_n \). Theoretically, since \( \varepsilon \) can be arbitrarily small, there is no problem to use the corresponding results for understanding of the large-sample behavior of \( R^2_n \). Indeed, the results discovered below show that the limiting distribution of \( R^2_n \) is a weighted sum of an infinite number of independently squared standard normal random variables with weights \( 1/\lambda_k \) decaying in a fashion proportional to \( 1/k^2 \). Practically, for any finite sample of size \( n \), the large-sample approximation to the distribution of \( R^2_n \) is valid as long as it can provide satisfactory numerical approximations. The use of \( \varepsilon = 1/[2(n + 1)] \) is suggested based on the fact that when mapped into the interval \((0, 1)\) as done in Section 3, the corresponding finite-sample extreme indices are \( 1/(n + 1) \) and \( n/(n + 1) = 1 - 1/(n + 1) \); See also Remark 6.1. Alternative values can be used and are discussed below.

The next critical step is to solve the eigensystem defined by the integral equation:

\[
f(t) = \lambda \int_{\varepsilon}^{1-\varepsilon} \sqrt{w(s)} \sqrt{w(t)} [\min(s, t) - st] f(s) ds.
\] (6.1)

It is easy to find that the solution satisfies the Sturm-Liouville equation (see, e.g., Anderson and Darling, 1952):

\[
h''(t) + \lambda \psi(t) h(t) = 0
\] (6.2)

where \( h(t) = f(t) \psi^{-\frac{1}{2}}(t) \). This is known as an eigenvalue problem. The solution can be found analytically and is summarized into the following theorem.

**Theorem 5.** The solution to the integral equation (6.1) is given by a sequence of \( \lambda_k = \omega^2_k + \frac{1}{4} \) with the corresponding eigenfunctions of two types. The first type is given by \( \omega_k s \) satisfying

\[
\tan \left( \omega_k \ln \frac{1 - \varepsilon}{\varepsilon} \right) = \frac{1}{2\omega_k}
\] (6.3)

and the corresponding eigenfunction

\[
f_k(x) \propto \frac{1}{\sqrt{t(1-t)}} \cos \left( \omega_k \ln \frac{t}{1-t} \right) \quad (t \in (\varepsilon, 1-\varepsilon)).
\]

The second type is given by \( \omega_k s \) satisfying

\[
\tan \left( \omega_k \ln \frac{1 - \varepsilon}{\varepsilon} \right) = -2\omega_k
\] (6.4)

and the corresponding eigenfunction

\[
f_k(x) \propto \frac{1}{\sqrt{t(1-t)}} \sin \left( \omega_k \ln \frac{t}{1-t} \right) \quad (t \in (\varepsilon, 1-\varepsilon)).
\]

Moreover, this solution corresponds to the Sturm-Liouville problem with the Sturm-Liouville equation (6.2) and the Robin boundary conditions

\[
f(\varepsilon) - 2(1-\varepsilon)f'(\varepsilon) = 0 \quad \text{and} \quad f(1-\varepsilon) + 2(1-\varepsilon)f'(1-\varepsilon) = 0
\]

and, thereby, all the eigenfunctions are orthogonal to each other.
Figure 5: An illustration of asymptotic approximation. The arrows indicate the asymptotic locations, where the \( z = \tan(\omega \ln \frac{1-\varepsilon}{\varepsilon}) \) curves intersect the curve \( z = \frac{1}{2\omega} \) and the line \( z = -2\omega \). The solid dot locations are locations of \( \omega \)'s computed via eigen-decomposition.

The proof of Theorem 5 is given in Appendix A.5. As depicted in Figure 5, the \( \omega_k \)'s satisfying (6.3) are in the intervals \( [k\pi, k\pi + \frac{1}{2}\pi) \), one in each interval, for \( k = 0, 1, 2, ... \), whereas the \( \omega_k \)'s satisfying (6.4) are in the intervals \( [k\pi - \frac{1}{2}\pi, k\pi) \), one in each interval, for \( k = 1, 2, ... \)

**Remark 6.1.** The normalizing constant \( C_n \) defined in (2.7) can be reset, if desirable, by making use of the following bounds for \( \sum_{i=1}^{n} \frac{1}{\lambda_i} \):

\[
\sum_{k=1}^{n} \frac{1}{\lambda_i} \approx \sum_{k=1}^{n} \frac{1}{4} + \left[ \frac{k\pi}{2\ln \frac{1-\varepsilon}{\varepsilon}} \right]^2 \geq \frac{4}{\pi} \ln \frac{1-\varepsilon}{\varepsilon} \int_{\ln \frac{1-\varepsilon}{\varepsilon}}^{\ln \frac{1+\varepsilon}{\varepsilon}} \frac{1}{1+t^2} \, dt \approx 2 \ln \frac{1-\varepsilon}{\varepsilon}
\]

and

\[
\sum_{k=1}^{n} \frac{1}{\lambda_i} \approx \sum_{k=1}^{n} \frac{1}{4} + \left[ \frac{k\pi}{2\ln \frac{1+\varepsilon}{\varepsilon}} \right]^2 \leq \frac{4}{\pi} \ln \frac{1-\varepsilon}{\varepsilon} \int_{0}^{\ln \frac{1+\varepsilon}{\varepsilon}} \frac{1}{1+t^2} \, dt \approx 2 \ln \frac{1-\varepsilon}{\varepsilon}.
\]

These approximations suggest in turn the use of \( \varepsilon \approx \frac{1}{2(n+1)} \).

Sort all eigenvalues obtained from the \( \omega_k \)s in (6.3) both (6.4) into \( \lambda_1 < \lambda_2 < ... \). Then, according
Figure 6: Performance of asymptotic approximation to the distribution of $R_n^2$ for $n = 10, 100,$ and 1000 obtained via a Monte Carlo approximation with 100,000 replicates. The three plots in the upper panel are the quantile-quantile plot in cubic-root scale with quantiles corresponding to the probabilities $i/1001$ for $i = 1, ..., 1000$, whereas the three plots in the lower panel are the corresponding probability-probability plot.

To Equation (4.5) of [Anderson and Darling (1952)], the asymptotic distribution of $R_n^2$ is that of

$$\sum_{i=1}^{\infty} \frac{X_i^2}{\lambda_i}$$

where $X_i^2$s are independently and identically distributed $\chi_i^2$ random variables. A simple Monte Carlo simulation-based study of evaluating this asymptotic distribution as an approximation to that of $R_n^2$ is summarized into Figure 6 using both quantile-quantile plots and probability-probability plots for $n = 10, 100,$ and 1,000. The Monte Carlo sample size used is 100,000; See [Davies (1980)] and [Duchesne and De Micheaux (2010)] for alternative computational methods. The truncated series $\sum_{i=1}^{n} \frac{X_i^2}{\lambda_i}$ was used, with $\varepsilon$ determined to match the $\lambda_1$. Such $\varepsilon$ values in all the three cases are close to $1/[2(n+1)]$. It is seen from these numerical results that the asymptotic approximation is satisfactory, even for small sample size.

### 6.2 Large-Sample Results for $\tilde{W}_n^2$ and $\tilde{R}_n^2$

#### 6.2.1 Gaussian process approximation and kernel matrices

Familiar asymptotic results for the $U_{(i)}$ process can be conveniently applied by writing:

$$\sqrt{n+2}(U_{(i)} - \mu_i) \approx B(t) = W(t) - tW(1) \text{ with } t = i/(n + 1),$$
where “≈” means that in the limit with $i/(n + 1) \to t$, the random variable on its left-hand side converges in distribution to the random variable on its right-hand side; See, e.g., [Anderson and Darling (1952)]. We can work with this Gaussian process approximation effectively as follows.

Define $Z_i = \sqrt{n + 1} [W(i/(n + 1)) - W((i - 1)/(n + 1))], i = 1, ..., n + 1$, and let $\bar{Z} = \frac{1}{n+1} \sum_{i=1}^{n+1} Z_i$. Then $Z_i$ are iid $N(0, 1)$. It follows that

$$
\mathcal{B}(i/(n + 1)) - \mathcal{B}((i - 1)/(n + 1)) = [W(i/(n + 1)) - W((i - 1)/(n + 1))] - [i/(n + 1) - (i - 1)/(n + 1)]W(1)
$$

$$
= [W(i/(n + 1)) - W((i - 1)/(n + 1))] - \frac{1}{n + 1} \sum_{j=1}^{n+1} [W(j/(n + 1)) - W((j - 1)/(n + 1))]
$$

$$
= (Z_i - \bar{Z})/\sqrt{n + 1}.
$$

Let

$$
C_n = \begin{bmatrix} 1 & -1/n + 1 \end{bmatrix},
$$

(6.6)
a symmetric $(n + 1) \times (n + 1)$ matrix and the centering operator for $Z = (Z_1, ..., Z_n, Z_{n+1})'$. Thus for all $c = 0, ..., n$,

$$
\sqrt{n + 2}[U_{(i)}^{(c)} - \mu_i] \approx \frac{1}{\sqrt{n + 1}} \tau_{c+[1:]} Z C_n Z
$$

(6.7)

where $Z = (Z_1, ..., Z_n, Z_{n+1})' \sim N_{n+1}(0, I)$ and $\tau_{c+[1:]}$ is the index vector for the elements of $Z$ at $c + 1, ..., c + i$, defined circularly. Let $A_i$ be the $(n + 1) \times (n + 1)$ matrix obtained by stacking these index vectors. That is, $A_k$ is circulant with its first row consisting of $k$ ones and $n + 1 - k$ zeros (in that order). So

$$
\sum_{c=0}^{n} (U_{(i)}^{(c)} - \mu_i)^2 \approx \frac{1}{(n + 1)(n + 2)} Z' C_n A_i' A_i C_n Z
$$

and, thereby,

$$
\sum_{c=0}^{n} \sum_{i=1}^{n} \psi_i (U_{(i)}^{(c)} - \mu_i)^2 \approx \frac{n + 1}{n + 2} Z' K_n(\psi) Z,
$$

(6.8)

where

$$
K_n(\psi) = \frac{1}{(n + 1)^2} \sum_{i=1}^{n} \psi_i C_n A_i' A_i C_n
$$

(6.9)

is the $(n + 1) \times (n + 1)$ kernel matrix with $\psi_i = w_i/[\mu_i(1 - \mu_i)]$ for $i = 1, ..., n$. To obtain the corresponding approximation to $C(U, w)$ using the above Gaussian process, we consider the following Taylor expansion of the $i$-th summand of Eq. (??) at $\mu_i$:

$$
-2 \left[ \mu_i \ln \frac{U_{(i)}}{\mu_i} + (1 - \mu_i) \ln \frac{1 - U_{(i)}}{1 - \mu_i} \right] \approx \frac{1}{\mu_i(1 - \mu_i)} (U_{(i)} - \mu_i)^2.
$$

This suggests the following approximation to the circularly symmetric test statistic $C(U, w)$:

$$
C(U, w) \approx \frac{1}{n + 2} Z' K_n(\psi) Z
$$

or, simply, $C(U, w) \approx \frac{1}{n} Z' K_n(\psi) Z$. 

15
The following theorem summarizes the properties regarding eigenvalues and eigenvectors of circulant matrices (Theorem 7 of Gray, 2006).

**Theorem 6.** Let $C$ be a $(n+1) \times (n+1)$ circulant matrix with its first row denoted by $c = (c_0, \ldots, c_n)'$, i.e., the $(k,j)$ entry of $C$ is given by $C_{k,j} = c_{(j-k) \mod (n+1)}$. Then $C$ has eigenvectors

$$
\nu(m) = \frac{1}{\sqrt{n+1}} (1, e^{-2\pi i m/(n+1)}, e^{-2\pi i 2m/(n+1)}, \ldots, e^{-2\pi i nm/(n+1)})', \quad m = 0, 1, \ldots, n,
$$

and corresponding eigenvalues

$$
\phi_m = \sum_{k=0}^{n} c_k e^{-2\pi i km/(n+1)}
$$

and can be expressed in the form $C = V \Phi V^*$, where $i$ is the unit imaginary number, $V$ has the eigenvectors as columns in order, the asterisk * denotes conjugate transpose, and $\Phi$ is diag($\phi_k$).

In particular all circulant matrices share the same eigenvectors, the same matrix $U$ works for all circulant matrices, and any matrix of the form $C = V \Phi V^*$ is circulant. Furthermore, for any two $(n+1) \times (n+1)$ circulant matrices $C$ and $B$, $C$ and $B$ commute, i.e., $CB = BC$, and $CB$, $\alpha C$, and $C + B$ are circulant matrices, where $\alpha$ is a scalar.
Since the eigenvalues of any real symmetric matrix are real, the symmetric circulant matrix $K_n(\psi)$ has $(n + 1)$ real eigenvalues

$$\phi_m = \sum_{k=0}^{n} c_k \cos \left( \frac{2\pi km}{n + 1} \right).$$

(6.12)

Moreover, a $(n + 1) \times (n + 1)$ symmetric circulant matrix $C$ satisfies the extra condition that $c_{n-i} = c_i$ and is thus determined by $[(n + 1)/2] + 1$ elements. The corresponding eigenvalues can be written as

$$\phi_m = c_0 + 2 \sum_{k=1}^{(n+1)/2-1} c_k \cos \left( \frac{2\pi km}{n + 1} \right) + c_{(n+1)/2} \cos (\pi km)$$

for $(n + 1)$ even, and

$$\phi_m = c_0 + 2 \sum_{k=1}^{n/2} c_k \cos \left( \frac{2\pi km}{n + 1} \right)$$

for $(n + 1)$ odd. These properties allow for efficient computation via fast discrete Fourier transform (Cooley and Tukey, 1965). The large-sample approximation to (6.11) for $\tilde{W}_n^2$ and $\tilde{R}_n^2$ is given in the next two subsections.

### 6.2.2 Circularized $W_n^2$, $\tilde{W}_n^2$

The continuum limit of the kernel structure (6.11) for the test statistic $W_n^2$ can be obtained and is summarized into the kernel function in the following theorem; See Appendix A.6 for the proof.

**Theorem 7.** If $w_k = 1$, i.e., $\psi_k \propto 1/|\mu_k(1 - \mu_k)|$, then in its continuum limit the kernel matrix $K_n(\psi)$ with elements (6.11) is given by

$$\kappa(t, s) = \begin{cases} 
2 \left[ (s - t) \ln(s - t) + [1 - (s - t)] \ln(1 - (s - t)) \right] + 1, & \text{if } t \leq s; \\
2 \left[ (t - s) \ln(t - s) + [1 - (t - s)] \ln(1 - (t - s)) \right] + 1, & \text{if } s < t.
\end{cases}$$

(6.13)

![Figure 7](image-url)

Figure 7: Quantile-Quantile plot of the large-sample approximation to the distribution of $\tilde{W}_n^2$ versus the true distribution in the cubic-root scale, obtained on 1,000,000 Monte Carlo samples. The quantile points are obtained for 1,000 equally spaced probabilities from $1/1001$ to $1 - 1/1001$. 

17
Remark 6.2. In case with \( w_k = 1 \), i.e., \( \psi_k = 1/[\mu_k(1 - \mu_k)] \), the diagonal elements of the kernel matrix are given by

\[
\sum_{k=1}^{n} \frac{k - \frac{k^2}{n+1}}{k(n + 1 - k)} = \frac{n}{n + 1}.
\]  

(6.14)

This implies that the sum of all the eigenvalues of the kernel matrix is \( n \). Because the rank of the circulant matrix \( C_n \) is \( n \) and the circulant matrix \( \sum_{k=1}^{n} \psi_k A_k' A_k \) is full rank, the kernel matrix is a rank-\( n \) matrix and thus has \( n \) non-zero eigenvalues with the zero eigenvalue given by

\[
\phi_0 = \sum_{k=0}^{n} c_k,
\]

where \( c = (c_0, \ldots, c_n) \) denotes the first row of the kernel matrix.

Remark 6.2 implies that \( E(C(U, w)) \approx 1 \) and, naturally, suggests \( \tilde{W}_n^2 \) defined in (4.3) for the preference of \( E(\tilde{W}_n^2) \approx 1 \). A numerical evaluation of the large sample-based approximation to the distribution of \( \tilde{W}_n^2 \) is shown by the quantile-quantile plots in Figure 7 for a selected cases of \( n = 10, 20, \) and 50. The quantile points are obtained for 1,000 equally spaced CDF values from 1/1001 to \( 1 - 1/1001 \) based on a Monte Carlo approximation of 1,000,000 replicates. The asymptotic kernel function (6.13) is used to compute \( n + 1 \) values as the first row of a corresponding kernel matrix:

\[
c_0 = \frac{1}{n + 1}, \quad c_k = 2 \left[ \frac{k_{n+1} \ln \left( \frac{k}{n+1} + \frac{n+1-k}{n+1} \right)}{n+1} + 1 \right], \quad k = 1, \ldots, n.
\]

One may see from Figure 7 that the approximation is satisfactory.

6.2.3 Circularized \( R_n^2, \tilde{R}_n^2 \)

The continuum limit of the kernel structure (6.11) for the test statistic \( R_n^2 \) can also be obtained and is summarized into the kernel function in the following theorem; See Appendix A.7 for the proof.

Theorem 8. If \( w_k = 1/[\mu_k(1 - \mu_k)] \), i.e., \( \psi_k \propto 1/[\mu_k(1 - \mu_k)]^2 \), then in its continuum limit the kernel matrix \( K_n(\psi) \) with elements (6.11) is given by

\[
\kappa(t, s) = \begin{cases} 
2 \left[ 2(s-t) - 1 \right] \ln \frac{s-t}{t-s} - 2, & \text{if } t \leq s; \\
2 \left[ 2(t-s) - 1 \right] \ln \frac{t-s}{t-s} - 2, & \text{if } s < t.
\end{cases}
\]

(6.15)

Remark 6.3. In case with \( w_k = 1/[\mu_k(1 - \mu_k)] \), i.e., \( \psi_k = 1/[\mu_k(1 - \mu_k)]^2 \), the diagonal elements of the kernel matrix are given by

\[
\frac{1}{(n+1)^2} \sum_{k=1}^{n} \psi_k C_n A_k' A_k C_n = (n+1)^2 \sum_{k=1}^{n} \frac{k - \frac{k^2}{n+1}}{[k(n + 1 - k)]^2} = 2 \sum_{k=1}^{n} \frac{1}{k}.
\]

(6.16)

This implies that the sum of all the eigenvalues of the kernel matrix is \( 2(n+1) \sum_{k=1}^{n} \frac{1}{k} \).

Remark 6.3 implies that \( E(B(U, w)) = 2 \sum_{k=1}^{n} \frac{1}{k} \) and, naturally, suggests the test statistic \( \tilde{R}_n^2 \) defined in (4.4) for the preference of \( E(\tilde{R}_n^2) \approx 1 \). A numerical evaluation of the large sample-based approximation to the distribution of \( \tilde{R}_n^2 \) was conducted in the same way as that for \( \tilde{W}_n^2 \) in the
previous subsection. The asymptotic kernel function $(6.15)$ is used to compute $n + 1$ values as the first row of a corresponding kernel matrix:

\[ c_0 = \frac{1}{2(n + 1) \sum_{k=1}^{n} \frac{1}{k}}, \]

\[ c_k = \frac{\left(\frac{2k}{n+1} - 1\right) \ln \frac{k}{n+1-k} - 1}{(n + 1) \sum_{k=1}^{n} \frac{1}{k}}, \quad k = 1, \ldots, n. \]

The results are displayed in Figure 8, which shows that the approximation is satisfactory, although this can be further improved by small sample corrections.

7 Concluding Remarks

Assessing goodness-of-fit is a fundamental problem in both applied and theoretical statistics in general, and in data-driven (or auto-)modeling in contemporary big data analysis in particular. This paper aimed to three goals toward both deep understanding of the problem and perfection of statistical methods. It provided a geometric intuition for understanding, which leads to the conclusion that $R_n^2$ can serve better than Anderson-Darling as an omnibus test. This is consistent with the discovery of Zhang (2002). Furthermore, it proposed the method of circularization and showed that circularized versions can have a better performance than their parent tests. In addition, this paper also established the asymptotic distributions of $R_n^2$ and the two circularly symmetric tests $\tilde{W}_n^2$ and $\tilde{R}_n^2$.

Performance of the proposed methods can also be investigated for distributions containing unknown parameters. This can be done with either the traditional approach, which replies on point estimations of the unknown parameter, or the inferential models approach of Martin and Liu (2015), which can be viewed as a generalized theory of the familiar method of pivotal quantity for constructing confidence intervals and hypothesis testing.
References

Anderson, T. W. and D. A. Darling (1952). Asymptotic theory of certain “goodness of fit” criteria based on stochastic processes. The annals of mathematical statistics, 193–212.

Anderson, T. W. and D. A. Darling (1954). A test of goodness of fit. Journal of the American statistical association 49(268), 765–769.

Cantoni, A. and P. Butler (1976a). Eigenvalues and eigenvectors of symmetric centrosymmetric matrices. Linear Algebra and its Applications 13(3), 275–288.

Cantoni, A. and P. Butler (1976b). Properties of the eigenvectors of persymmetric matrices with applications to communication theory. IEEE Transactions on Communications 24(8), 804–809.

Carmen Pardo, M., Y. Lu, and A. M. Franco-Pereira (2022). Extensions of empirical likelihood and chi-squared-based tests for ordered alternatives. Journal of Applied Statistics 49(1), 24–43.

Cooley, J. W. and J. W. Tukey (1965). An algorithm for the machine calculation of complex fourier series. Mathematics of computation 19(90), 297–301.

Cramér, H. (1928). On the composition of elementary errors. Scandinavian Actuarial Journal 1, 13–74.

David, H. A. and H. N. Nagaraja (2004). Order statistics. John Wiley & Sons.

Davies, R. B. (1980). Algorithm as 155: The distribution of a linear combination of $\chi^2$ random variables. Applied Statistics, 323–333.

Duchesne, P. and P. L. De Micheaux (2010). Computing the distribution of quadratic forms: Further comparisons between the liu–tang–zhang approximation and exact methods. Computational Statistics & Data Analysis 54(4), 858–862.

Gray, R. M. (2006). Toeplitz and circulant matrices: A review. Foundations and Trends® in Communications and Information Theory 2(3), 155–239.

Kolmogorov, A. (1933). Sulla determinazione empirica di una legge di distribuzione. G. Ist. Ital. Attuari. 4, 83–91.

Lehmann, E. L. and J. P. Romano (2005). Testing statistical hypotheses, Volume 3. Springer.

Makhoul, J. (1981). On the eigenvectors of symmetric toeplitz matrices. IEEE Transactions on Acoustics, Speech, and Signal Processing 29(4), 868–872.

Martin, R. and C. Liu (2015). Inferential models: reasoning with uncertainty, Volume 145. CRC Press.

Pearson, K. (1900). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science 50(302), 157–175.

Shapiro, S. S. and M. B. Wilk (1965). An analysis of variance test for normality (complete samples). Biometrika 52(3/4), 591–611.
Sinclair, C. and B. Spurr (1988). Approximations to the distribution function of the anderson—darling test statistic. Journal of the American Statistical Association 83(404), 1190–1191.

Smirnov, N. V. (1939). On the deviation of the empirical distribution function. Rec. Math.[Mathematicheskii Sbornik] NS 6, 3–26.

Stephens, M. A. (1974). Edf statistics for goodness of fit and some comparisons. Journal of the American statistical Association 69(347), 730–737.

von Mises, R. E. (1928). Wahrscheinlichkeit. Statistik und Wahrheit.

Wolfram Research Inc., M. (2022). Mathematica, Version 13.1. Champaign, IL, 2022.

Zhang, J. (2002). Powerful goodness-of-fit tests based on the likelihood ratio. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 64(2), 281–294.

Zhang, J. (2010). Statistical inference with weak beliefs. Ph.D. thesis, Purdue University, West Lafayette, IN.

Zolotarev, V. M. (1961). Concerning a certain probability problem. Theory of Probability and Its Applications 6, 201–204.
A Proofs of Theorems

A.1 Proof of Theorem 1

Using the popular technique for deriving the expectation of $\ln(X)$ when $X$ is a Beta random variable, we have

$$E[\ln(U(i))] = \frac{1}{\Beta(\alpha, \beta)} \frac{\partial}{\partial \alpha} \int_0^1 u^{\alpha-1}(1-u)^{\beta-1} du \bigg|_{\alpha=i, \beta=n+1-i} = \psi(i) - \psi(n + 1).$$

The claimed results can be verified using such standard techniques with tedious algebraic operations.

A.2 Proof of Theorem 2

From the pdf of the joint distribution $U(i)$ and $U(j)$, we have for all $k$ and $l$,

$$E\left\{U_k(i)[1 - U(j)]^l\right\} = \frac{(i + k - 1)! (n - j + l)!}{(i - 1)! (n - j)!} \frac{n!}{(n + k + l)!}.$$ 

With this identify and tedious routine algebraic operations, one can verify the claimed results.

A.3 Proof of Theorem 3

Let $B(i/(n+1)) = \sqrt{n+2}(U(i) - E(U(i)))$ for $i = 1, ..., n$. As $n \to \infty$, $B(i)$ converges in distribution to a Brownian bridge. The results on the covariance structure of the Brownian bridge are well-known and easy-to-prove. So, our proof here will focus on the results on $\Cov(B^2(s), B^2(t))$.

Write the Brownian bridge using the Brownian motion $W(t)$, $t \in [0, 1]$ as follows

$$B(t) = W(t) - tW(1)$$

$$E(B^2(t)) = E([(1 - t)W(t) - t(W(1) - W(t))]^2) = t(1 - t)$$

For $0 < s < t < 1$, it is easy to see that

$$E\left[W^2(s)W^2(t)\right] = 3s^2 + s(t - s),$$

and

$$E\left[B^2(s)B^2(t)\right] = 2s^2 + st - 5s^2t - st^2 + 3s^2t^2.$$ 

Thus, the result follows.

A.4 Proof of Theorem 4

The continuum limit of the Lagrange auxiliary equation is obtained from

$$2(1 - s)^2 \int_0^s t^2 \psi(t) dt + 2s^2 \int_s^1 (1 - t)^2 \psi(t) dt \propto 1,$$

a constant for all $s \in [\varepsilon, 1 - \varepsilon]$ and $\varepsilon \in (0, 1/2)$, and is given by

$$- 4(1 - s) \int_0^s t^2 \psi(t) dt + 4s \int_s^1 (1 - t)^2 \psi(t) dt = 0 \quad (A.1)$$
because $2(1 - s)^2s^2\psi(s) - 2s^2(1 - s)^2\psi(s) = 0$ for all $s \in [\epsilon, 1 - \epsilon]$. Equation (A.1) is an integral equation, known as the Fredholm equation, which does not have a general solution. Here, we solve it by converting it into a differential equation.

Differentiate the both sides of Equation (A.1) to obtain

$$\int_0^s t^2 \psi(t) dt + \int_1^s (1 - t)^2 \psi(t) dt - s(1 - s) \psi(s) = 0$$

for some positive constant $c_0$. Equation (A.2)

Differentiating the both sides of Equation (A.2), we obtain

$$\psi'(s) - 2 \left( \frac{1}{1 - s} - \frac{1}{s} \right) \psi(s) = 0$$

Applying the method of separation of variables, we get the solution to Equation (A.3):

$$\psi(t) = \frac{c_0}{t^2 (1 - t)^2} \quad (t \in [\epsilon, 1 - \epsilon]),$$

for some positive constant $c_0$.

Note that $\psi(.)$ on $(0, \epsilon)$ and $(1 - \epsilon, 1)$ must satisfy the condition (A.1):

$$-4(1 - s) \int_0^\epsilon t^2 \psi(t) dt + 4s \int_1^{1-\epsilon} (1 - t)^2 \psi(t) dt - 4(1 - s) \int_\epsilon^s \frac{c_0}{(1 - t)^2} dt + 4s \int_s^{1-\epsilon} \frac{c_0}{t^2} dt = 0$$

for all $s \in [\epsilon, 1 - \epsilon]$. We need to show that such a $\psi(.)$ exists. Taking $\psi(t) = \frac{c_0 h(\epsilon)}{t^2(1-t)^2}$ with $h(\epsilon) = 1/\epsilon$, for example, we have for the left hand side of (A.5):

$$4[s - (1 - s)] \int_0^\epsilon \frac{c_0 h(\epsilon)}{(1 - t)^2} dt - 4(1 - s) \left[ \frac{c_0}{1-t} \bigg|_{\epsilon}^{s} \right] + 4s \left[ - \frac{c_0}{t} \bigg|_{s}^{1-\epsilon} \right]$$

$$= 4(2s - 1)c_0 h(\epsilon) \frac{\epsilon}{1 - \epsilon} + 4(1 - 2s)c_0 \frac{1}{1 - \epsilon}$$

Now, letting $\epsilon \to 0$, we obtain from (A.4) that

$$\psi(t) = \frac{c_0}{t^2 (1 - t)^2} \quad (t \in (0, 1)),$$

completing the proof.

### A.5 Proof of Theorem 5

Let $h(t) = f(t)\psi^{-1/2}(t)$. Recall from Anderson and Darling (1952) that

$$h''(t) + \lambda \psi(t) h(t) = 0.$$

In the proof, we work with the following transformation

$$x = 2t - 1 \quad \text{and, thereby,} \quad t = (1 + x)/2, \quad (x \in [-1 + 2\epsilon, 1 - 2\epsilon]).$$

So have $\psi(t = (1 + x)/2) = 16/(1 - x^2)^2$ and $g(x) = f(t = (1 + x)/2)\psi^{-1/2}(t = (1 + x)/2) = \frac{1}{4}(1 - x^2)f(t = (1 + x)/2) = \frac{1}{4}(1 - x^2)\tilde{f}(x)$. Thus,

$$g'(x) = \frac{1}{4} \left[ (1 - x^2)\tilde{f}'(x) - 2x\tilde{f}(x) \right]$$
and
\[ g''(x) = \frac{1}{4} \left[ (1 - x^2) \ddot{f}(x) - 4x \dot{f}(x) - 2 \ddot{f}(x) \right]. \]

It follows from
\[ g''(x) + \frac{4\lambda}{(1 - x^2)^2} g(x) = 0 \quad (A.7) \]
that
\[ (1 - x^2) \ddot{f}(x) - 4x \dot{f}(x) + \left[ \frac{4\lambda}{1 - x^2} - 2 \right] \dot{f}(x) = 0 \quad (A.8) \]

The second-order differential equation (A.8) can be solved with Mathematica (Wolfram Research Inc., 2022) or the trial solution method with
\[ y(x) = c(1 - x^2)^{-\tau} e^{\xi \arctanh(x)} \]
where \( c, \tau, \) and \( \xi \) are constant, \( \arctanh(.) \) is the inverse of the hyperbolic function \( \tanh: \)
\[ \arctanh(x) = \frac{1}{2} \ln \frac{1 + x}{1 - x} \quad (x \in (-1, 1)). \]

The general solution is given by
\[ y(x) = c_1 (1 - x^2)^{-\frac{1}{2}} e^{\xi_1 \arctanh(x)} + c_2 (1 - x^2)^{-\frac{1}{2}} e^{\xi_2 \arctanh(x)} \quad (A.9) \]
where \( \xi_1 \) and \( \xi_2 \) are the two roots of the quadratic function
\[ \xi^2 + 4\lambda - 1 = 0. \]

Incidentally, it is easy to see that the finite-sample counterpart covariance matrix is \textit{doubly symmetric}, \textit{i.e.}, symmetric about both the main diagonal and the secondary diagonal; See Makhoul (1981), Cantoni and Butler (1976b), and Cantoni and Butler (1976a) for more details.

If \( \lambda \leq 1/4 \), then the general solution \( y(x) \) is given as
\[ y(x) = (1 - x^2)^{-\frac{1}{2}} \left[ c_1 e^{-\theta \arctanh(x)} + c_2 e^{\theta \arctanh(x)} \right] \]
where \( \theta = \sqrt{1 - 4\lambda} \). Making use of the following basic calculus results
\[
\begin{align*}
\int (1 - x^2)^{-\frac{3}{2}} e^{\theta \arctanh(x)} dx &= -\frac{(\theta - x)(1 - x^2)^{-\frac{1}{2}} e^{\theta \arctanh(x)}}{1 - \theta^2} + C, \\
\int x(1 - x^2)^{-\frac{3}{2}} e^{\theta \arctanh(x)} dx &= -\frac{(\theta x - 1)(1 - x^2)^{-\frac{1}{2}} e^{\theta \arctanh(x)}}{1 - \theta^2} + C, \\
\int (1 - x^2)^{-\frac{3}{2}} e^{-\theta \arctanh(x)} dx &= \frac{(\theta + x)(1 - x^2)^{-\frac{1}{2}} e^{-\theta \arctanh(x)}}{1 - \theta^2} + C, \\
\int x(1 - x^2)^{-\frac{3}{2}} e^{-\theta \arctanh(x)} dx &= \frac{(\theta x + 1)(1 - x^2)^{-\frac{1}{2}} e^{-\theta \arctanh(x)}}{1 - \theta^2} + C,
\end{align*}
\]
and
\[
\begin{align*}
\int x(1 - x^2)^{-\frac{3}{2}} e^{\theta \arctanh(x)} dx &= \frac{(\theta x + 1)(1 - x^2)^{-\frac{1}{2}} e^{\theta \arctanh(x)}}{1 - \theta^2} + C,
\end{align*}
\]
where \( C \) is a constant, we can see that non-trivial solutions require that
\[ 2\epsilon(1 + \theta) e^{\theta \arctanh(1 - 2\epsilon)} + 2\epsilon(1 - \theta) e^{-\theta \arctanh(1 - 2\epsilon)} = 0. \]
Since $0 \leq \theta \leq 1$, there are no non-trivial solution if $\lambda \leq 1/4$.

If $\lambda > 1/4$, then routine algebraic operations on complex numbers lead to the general solution $y(x)$ given as
\[
y(x) = \frac{1}{(1 - x^2)^{3/2}} \left[ c_1 \cos(\theta \arctanh(x)) + c_2 \sin(\theta \arctanh(x)) \right],
\] (A.10)
where $\theta = 2\omega = \sqrt{4\lambda - 1}$.

To find solutions satisfying the integral equation (6.1), we can make use of the following indefinite integrals
\[
\int (1 - x^2)^{-3/2} \sin(\theta \arctanh(x)) dx = \frac{-\theta \cos(\theta \arctanh(x)) + x \sin(\theta \arctanh(x))}{(1 + \theta^2)\sqrt{1 - x^2}} + C,
\]
\[
\int (1 - x^2)^{-3/2} \cos(\theta \arctanh(x)) dx = \frac{x \cos(\theta \arctanh(x)) + \theta \sin(\theta \arctanh(x))}{(1 + \theta^2)\sqrt{1 - x^2}} + C,
\]
\[
\int x(1 - x^2)^{-3/2} \sin(\theta \arctanh(x)) dx = \frac{-\theta x \cos(\theta \arctanh(x)) + \sin(\theta \arctanh(x))}{(1 + \theta^2)\sqrt{1 - x^2}} + C,
\]
and
\[
\int x(1 - x^2)^{-3/2} \cos(\theta \arctanh(x)) dx = \frac{x \cos(\theta \arctanh(x)) + \theta \sin(\theta \arctanh(x))}{(1 + \theta^2)\sqrt{1 - x^2}} + C,
\]
where $C$ is a constant.

Equating $f(t)$ and $\lambda \int_0^1 (\min(t, s) - ts) \sqrt{\psi(t)\sqrt{\psi(s)}f(s)} ds$, with standard calculus operations, we can obtain
\[c_1 A(x) + c_2 B(x) = 0,
\]
where, omitting tedious details of derivation,
\[
A(x) = \frac{-\varepsilon \cos(\theta \arctanh(1 - 2\varepsilon)) + \varepsilon \theta \sin(\theta \arctanh(1 - 2\varepsilon))}{\sqrt{\varepsilon(1 - \varepsilon)}} \frac{1}{1 - x^2}
\]
and
\[
B(x) = \frac{-\varepsilon \theta \cos(\theta \arctanh(1 - 2\varepsilon)) + \varepsilon \sin(\theta \arctanh(1 - 2\varepsilon))}{\sqrt{\varepsilon(1 - \varepsilon)}} \frac{x}{1 - x^2}.
\]

The non-trivial solutions require that with $c_1^2 + c_2^2 \neq 0$, $c_1 A(x) + c_2 B(x) = 0$ for all $x \in [-1 + 2\varepsilon, 1 - 2\varepsilon]$. This amounts to requiring
\[
\cos(\theta \arctanh(1 - 2\varepsilon)) - \theta \sin(\theta \arctanh(1 - 2\varepsilon)) = 0
\] (A.11)
for $A(x) = 0$, that is,
\[
\tan \left( \omega \ln \frac{1 - \varepsilon}{\varepsilon} \right) = \frac{1}{2\omega},
\] (A.12)
and
\[
\theta \cos(\theta \arctanh(1 - 2\varepsilon)) + \sin(\theta \arctanh(1 - 2\varepsilon)) = 0
\] (A.13)
for $B(x) = 0$, that is,
\[
\tan \left( \omega \ln \frac{1 - \varepsilon}{\varepsilon} \right) = -2\omega.
\] (A.14)

It is easy to see that the claimed results follow (A.12) and (A.14).
Regarding the claim on the presentation of the eigenvalue problem as a Sturm-Liouville problem, we prove it by establishing Robin boundary conditions for the fundamental initial conditions (A.11) and (A.13). For notational convenience, here we take \( a = -1 + 2\epsilon \) and \( b = 1 - 2\epsilon \). Differentiate (A.10) to obtain

\[
y'(x) = c_1 (1 - x^2)^{-\frac{3}{2}} [x \cos (\theta \arctanh(x)) - \theta \sin (\theta \arctanh(x))] \\
+ c_2 (1 - x^2)^{-\frac{3}{2}} [x \sin (\theta \arctanh(x)) + \theta \cos (\theta \arctanh(x))],
\]

which implies that

\[
[4\epsilon(1 - \epsilon)]^\frac{3}{2} y'(a) = c_1 \left[ (-1 + 2\epsilon) \cos (\theta \arctanh(-1 + 2\epsilon)) - \theta \sin (\theta \arctanh(-1 + 2\epsilon)) \right] \\
+ c_2 \left[ (-1 + 2\epsilon) \sin (\theta \arctanh(-1 + 2\epsilon)) + \theta \cos (\theta \arctanh(-1 + 2\epsilon)) \right]
\]

and

\[
[4\epsilon(1 - \epsilon)]^\frac{3}{2} y'(b) = c_1 \left[ (1 - 2\epsilon) \cos (\theta \arctanh(1 - 2\epsilon)) - \theta \sin (\theta \arctanh(1 - 2\epsilon)) \right] \\
+ c_2 \left[ (1 - 2\epsilon) \sin (\theta \arctanh(1 - 2\epsilon)) + \theta \cos (\theta \arctanh(1 - 2\epsilon)) \right]
\]

Also, the values of \( y(x) \) at the two end points are obtained from (A.10) as

\[
[4\epsilon(1 - \epsilon)]^\frac{1}{2} y(a) = c_1 \cos (\theta \arctanh(-1 + 2\epsilon)) + c_2 \sin (\theta \arctanh(-1 + 2\epsilon))
\]

and

\[
[4\epsilon(1 - \epsilon)]^\frac{1}{2} y(b) = c_1 \cos (\theta \arctanh(1 - 2\epsilon)) + c_2 \sin (\theta \arctanh(1 - 2\epsilon))
\]

Consider the following equivalent Robin boundary conditions

\[
\alpha_1 [4\epsilon(1 - \epsilon)]^\frac{1}{2} y(a) + [4\epsilon(1 - \epsilon)]^\frac{3}{2} y'(a) = 0 \\
\alpha_2 [4\epsilon(1 - \epsilon)]^\frac{1}{2} y(b) + [4\epsilon(1 - \epsilon)]^\frac{3}{2} y'(b) = 0,
\]

that is,

\[
0 = c_1 \left[ (\alpha_1 - 1 + 2\epsilon) \cos (\theta \arctanh(1 - 2\epsilon)) + \theta \sin (\theta \arctanh(1 - 2\epsilon)) \right] \\
+ c_2 \left[ (-\alpha_1 - 1 + 2\epsilon) \sin (\theta \arctanh(1 - 2\epsilon)) + \theta \cos (\theta \arctanh(1 - 2\epsilon)) \right] \\
0 = c_1 \left[ (\alpha_2 + 1 - 2\epsilon) \cos (\theta \arctanh(1 - 2\epsilon)) - \theta \sin (\theta \arctanh(1 - 2\epsilon)) \right] \\
+ c_2 \left[ (\alpha_2 + 1 - 2\epsilon) \sin (\theta \arctanh(1 - 2\epsilon)) + \theta \cos (\theta \arctanh(1 - 2\epsilon)) \right]
\]

For non-trivial solutions of \( y(x) \), the determinant of the matrix of the coefficients in the above system of two linear equations must be zero. With routine algebraic operations, it is easy to see that this is satisfied if and only if \( \alpha_1 = -2\epsilon \) and \( \alpha_2 = 2\epsilon \). This leads to the Robin boundary conditions:

\[
y(a) - 2(1 - \epsilon)y'(a) = 0 \text{ and } y(b) + 2(1 - \epsilon)y'(b) = 0.
\]

From the Sturm-Liouville theory on the orthogonality of the solutions to (A.7) that satisfy Robin boundary conditions, it is known that

\[
\int_{-1+2\epsilon}^{1-2\epsilon} \frac{1}{(1-x^2)^2} g_1(x)g_2(x)dx = 0 \quad \text{(A.15)}
\]

holds for any two different solutions \( g_1(x) \) and \( g_2(x) \). Note that \( g(x) = f(t)\psi^{-\frac{1}{2}}(t) \) with \( t = (1 + x)/2 \). Equation (A.16) can be written as

\[
\frac{1}{2} \int_{-1+2\epsilon}^{1-2\epsilon} \frac{1}{4[t(1-t)]^2} \psi^{-1}(t)dt = \frac{1}{8} \int_{-1+2\epsilon}^{1-2\epsilon} f_1(t)f_2(t)dt = 0 \quad \text{(A.16)}
\]

This completes the proof.
A.6 Proof of Theorem 7

Due to the symmetry property, it is sufficient to consider the case $s \geq t$. In this case, the element in the continuum limit with $i/(n + 1) \to t$ and $j/(n + 1) \to s$, where $0 < t < s < 1$, is obtained as follows: Note that $n + 1 - (j - i) > 0$,

$$
\lim_{n \to \infty} \frac{1}{(n + 1)^2} \left[ \sum_{k=j-i}^{n} \psi_k[k - (j - i)] + \sum_{k=n+1-(j-i)}^{n} \psi_k[k - i + j - (n + 1)] - \sum_{k=1}^{n} \psi_k k^2 \right]
$$

$$
\lim_{n \to \infty} \left[ \sum_{k=j-i}^{n} \frac{k - (j - i)}{k(n + 1 - k)} + \sum_{k=n+1-(j-i)}^{n} \frac{k - i + j - (n + 1)}{k(n + 1 - k)} - \frac{1}{n + 1} \sum_{k=1}^{n} \frac{k}{n + 1 - k} \right]
$$

$$
\lim_{n \to \infty} \left[ -\frac{j - i}{n + 1} \sum_{k=1}^{n+1-(j-i)} \frac{1}{k} - \frac{j - i}{n + 1} \sum_{k=j-i}^{n} \frac{1}{k} + \frac{j - i}{n + 1} \sum_{k=1}^{n} \frac{1}{k} \right]
$$

\[ (A.17) \]

If $j - i \geq (n + 1)/2$, that is, $j - i \geq n + 1 - (j - i)$ and $s - t \geq \frac{1}{2}$, then (A.17) becomes

$$
\lim_{n \to \infty} \left[ \frac{2(j - i)}{n + 1} \sum_{k=n+1-(j-i)+1}^{j-i} \frac{1}{k} + \frac{(j - i)}{n + 1 - (j - i)} \right] - \frac{1}{n + 1 - (j - i)} \sum_{k=n+1-(j-i)+1}^{n} \frac{1}{k} + \frac{n}{n + 1} \right]
$$

\[ (A.17) \]

where the method of series estimation with integrals is used. If $j - i < (n + 1)/2$, that is, $j - i < n + 1 - (j - i)$ and $s - t < \frac{1}{2}$, then (A.17) becomes

$$
\lim_{n \to \infty} \left[ \frac{-j - i}{n + 1} \sum_{k=j-i+1}^{j-i} \frac{1}{k} - \frac{j - i}{n + 1} \sum_{k=j-i}^{n+1-(j-i)-1} \frac{1}{k} + \frac{j - i}{n + 1} \sum_{k=1}^{n} \frac{1}{k} \right]
$$

\[ (A.17) \]

where the method of series estimation with integrals is used again. For the $s = t = i/(n + 1)$ case, it is easy to see that a simplified version of the above derivation gives the claimed result. To summarize, we have Equation (6.13) and complete the proof.
A.7 Proof of Theorem 8

Due to the symmetry property, it is sufficient to consider the case \( s \geq t \). In this, the element in the continuum limit with \( t = i/(n+1) \) and \( s = j/(n+1) \), where \( 0 < t < s < 1 \), is obtained as follows:

\[
\lim_{n \to \infty} \frac{1}{(n+1)^2} \left[ \sum_{k=j-i}^{n} \psi_k[k-(j-i)] + \sum_{k=n+1-(j-i)}^{n} \psi_k[k-i+j-(n+1)] - \sum_{k=1}^{n} \psi_k \frac{k^2}{n+1} \right]
= \lim_{n \to \infty} \frac{1}{(n+1)^2} \left[ \sum_{k=j-i}^{n} \frac{k-(j-i)}{k(n+1-k)^2} + \sum_{k=n+1-(j-i)}^{n} \frac{k-i+j-(n+1)}{k(n+1-k)^2} - \frac{1}{n+1} \sum_{k=1}^{n} \frac{1}{(n+1-k)^2} \right]
= \lim_{n \to \infty} \frac{1}{(n+1)^2} \sum_{\ell \in \{j-i,n+1-(j-i)\}} \sum_{k=\ell}^{n} \frac{1}{n+1-k} \left[ \frac{1}{k} + \frac{1}{n+1-k} \right] - \frac{j-i}{n+1} \sum_{k=j-i}^{n} \left[ \frac{1}{k} + \frac{1}{n+1-k} \right]^2 - \sum_{k=1}^{n} \frac{1}{k^2} \right]. \tag{A.18}
\]

If \( j-i < (n+1)/2 \), that is, \( j-i < n+1-(j-i) \) and \( s-t < \frac{1}{2} \), then (A.18) becomes

\[
\lim_{n \to \infty} \frac{1}{(n+1)^2} \sum_{k=j-i}^{n} \frac{1}{n+1-k} - \frac{j-i}{n+1} \sum_{k=j-i}^{n} \left[ \frac{1}{k} + \frac{1}{n+1-k} \right]^2 - \sum_{n=1}^{n} \frac{1}{k^2} \] \[
= 2 \ln \frac{1-(s-t)}{s-t} - (s-t) \int_{s-t}^{1-(s-t)} \left( \frac{1}{x} + \frac{1}{1-x} \right)^2 \, dx - 2 \int_{s-t}^{1} \frac{1}{x^2} \, dx
= 4 \left[ (s-t) - \frac{1}{2} \right] \ln \frac{s-t}{1-(s-t)} - 2,
\]

where the method of series estimation with integrals is used. If \( j-i \geq (n+1)/2 \), that is, \( j-i \geq n+1-(j-i) \) and \( s-t \geq \frac{1}{2} \), then (A.18) becomes

\[
\lim_{n \to \infty} \frac{1}{(n+1)^2} \left[ - \frac{2}{n+1} \sum_{k=n+1-(j-i)+1}^{j-i} \frac{1}{n+1-k} - 2 \sum_{k=1}^{j-i} \frac{1}{(n+1-k)^2} + \frac{j-i}{n+1} \sum_{k=n+1-(j-i)}^{j-i-1} \left( \frac{1}{k} + \frac{1}{n+1-k} \right)^2 \right]
= -2 \ln \frac{s-t}{1-(s-t)} - 2 \int_{0}^{s-t} \frac{1}{(1-x)^2} \, dx + (s-t) \int_{1-(s-t)}^{s-t} \left( \frac{1}{x} + \frac{1}{1-x} \right) \, dx
= 4 \left[ (s-t) - \frac{1}{2} \right] \ln \frac{s-t}{1-(s-t)} - 2,
\]

the same as that in the \( s-t < \frac{1}{2} \) case, where the method of series estimation with integrals is used again. For the \( s=t=i/(n+1) \) case, it is easy to see that a simplified version of the above derivation gives the claimed result. To summarize, we have Equation (6.15) and complete the proof.