Statistical Papers (2025) 66:135
https://doi.org/10.1007/s00362-025-01751-5

REGULAR ARTICLE

Efficient eigenvalue approximation in covariance operators via Rayleigh–Ritz with statistical applications

Bruno Ebner[1] · **M. Dolores Jiménez-Gamero**[2] · **Bojana Milošević**[3]

Received: 29 January 2025 / Revised: 8 July 2025

Abstract

Finding the eigenvalues connected to the covariance operator of a centered Hilbert-space valued Gaussian process is genuinely considered a hard problem in several mathematical disciplines. In Statistics this problem arises for instance in the asymptotic null distribution of goodness-of-fit test statistics of weighted L^2-type as well as in the limit distribution of degenerate U-statistics. For this problem we present the Rayleigh–Ritz method to approximate the eigenvalues. The usefulness of these approximations is shown by high lightening implications such as critical value approximation and theoretical comparison of test statistics by means of Bahadur efficiencies.

Keywords Covariance operator · Eigenvalues · Rayleigh-Ritz method · Gaussian Processes · Statistics

✉ Bruno Ebner
Bruno.Ebner@kit.edu

M. Dolores Jiménez-Gamero
dolores@us.es

Bojana Milošević
bojana@matf.bg.ac.rs

1 Institute of Stochastics, Karlsruhe Institute of Technology (KIT), Englerstr. 2, 76133 Karlsruhe, Germany

2 Dpto. Estadística e Investigación Operativa, Universidad de Sevilla, Avda Reina Mercedes sn, 41012 Sevilla, Spain

3 Faculty of Mathematics, University of Belgrade, Studentski trg 16, Belgrade, Serbia

Published online: 20 September 2025

1 Introduction

Weighted L^2-statistics are ubiquitous in goodness-of-fit and time series testing problems, see Baringhaus et al. (2017) for a comprehensive list of articles up to 2015 and for more recent advancements, see Allison et al. (2023); Allison and Pretorius (2017); Bahraoui et al. (2018); Betsch and Ebner (2019); Hadjicosta and Richards (2020a); Henze and Jiménez-Gamero (2019); Henze et al. (2019); Henze and Mayer (2020); Hušková et al. (2019). The limit null distribution of weighted L^2-statistics is typically that of $\|Z\|^2$, where Z is a Gaussian process taking values in a suitable function space and $\|\cdot\|$ is the corresponding norm. By the Karhunen-Loève expansion, the distribution of $\|Z\|^2$ is that of $\sum_{j=1}^{\infty} \lambda_j N_j^2$, where $N_1, N_2, \ldots$ are independent identically distributed (iid) standard normal random variables, and $\{\lambda_j\}_{j\in\mathbb{N}}$ is the sequence of positive eigenvalues of the covariance operator of Z. While the covariance kernels of the Gaussian process are often explicitly known, their complexity typically halts further analytical calculations. Hence, critical values for these tests are commonly determined via Monte Carlo simulation. In a few special cases, the pertaining eigenvalues can be calculated analytically (Baringhaus et al. 2018; Baringhaus and Henze 2008; Bilodeau and de Micheaux 2005; Ebner 2023; Henze and Nikitin 2000; Matsui and Takemura 2005), particularly in relation to the classic eigenvalue problem of the Brownian bridge covariance kernel. For a rigorous derivation of the solution, refer to Deheuvels and Martynov (2003); Shorack and Wellner (1986). Other approaches to approximate the eigenvalues include using the quadrature method (Bilodeau and de Micheaux 2005), numerically finding the roots of the associated Fredholm determinant (Ebner and Henze 2023; Stephens 1976, 1977), employing a stochastic Monte Carlo-type approach (Ebner and Henze 2023; Fan et al. 2017), or approximating the kernel (Božin et al. 2020; Jiménez-Gamero et al. 2015; Novoa-Muñoz and Jiménez-Gamero 2016). Note that the same structure of limit distribution connected to the covariance operator of a centered Gaussian process occurs in the case of degenerate U-statistics. Recent examples are found in e.g. Aleksić and Milošević (2024); Allison et al. (2022); Bose et al. (2023); Cuparić et al. (2022a); Milošević (2020); Shi et al. (2022).

The primary objective of this paper is to introduce and apply the Rayleigh–Ritz method—previously overlooked in statistical literature—as a novel approach to approximating the largest m eigenvalues $\lambda_1, \ldots, \lambda_m$ of the covariance kernel, as discussed in Kanwal (1971), Section 7.9. This method presents a robust alternative for cases where traditional techniques encounter limitations due to computational complexity. The secondary aim is to address gaps in the literature by providing accurate approximations of these eigenvalues. The third objective is to present two illustrative applications in which knowledge of the eigenvalues yields deeper insights into the test statistics under consideration.

To achieve this, the paper is organized as follows: Section 2 presents the eigenvalue problem; Section 3 details the Rayleigh–Ritz method; Section 4 applies the method to both established and novel eigenvalue problems in the literature for various supports. In Sect. 5 we examine the eigenvalue problem for covariance kernels depending on unknown parameters. Finally, Sect. 6 discusses the implications of knowing these eigenvalues, including a method for approximating the quantiles of the asymptotic distribution and providing approximate Bahadur efficiencies for several examples.

2 The eigenvalue problem

Let $\mathcal{B}^d$ denote the Borel σ-algebra of subsets of $\mathbb{R}^d$, with $M \neq \emptyset$ being an element of $\mathcal{B}^d$ and μ a finite measure on $\mathcal{B}^d_M = \mathcal{B}^d \cap M$. Define $L^2 = L^2(M, \mathcal{B}^d_M, \mu)$ as the (separable) Hilbert space consisting of (equivalence classes of) square-integrable measurable functions on *M*, equipped with the inner product $\langle g, h \rangle = \int_M gh \, \mathrm{d}\mu$. Moreover, let $\|h\|^2_{L^2} = \int_M h^2 \, \mathrm{d}\mu$ and let $Z = (Z(t), t \in M)$ be a centered Gaussian process that can be considered a random element of L^2 with $\mathbb{E}\|Z\|^2_{L^2} < \infty$. The distribution of *Z* is uniquely characterized by its covariance kernel *K*(*s*, *t*), $s, t \in M$. Consequently, as is well-known [see Duan and Wang (2014), Chapter 3], the distribution of $\|Z\|^2_{L^2}$ corresponds to $\sum_{j=1}^\infty \lambda_j N_j^2$, where N_j are iid standard normal random variables and λ_j are the positive eigenvalues, at most countable, associated with eigenfunctions f_j of the (linear second-order homogeneous Fredholm) integral equation

$$\lambda f(s) = \mathcal{K} f(s), \quad s \in M, \tag{1}$$

where $\mathcal{K} : L^2 \mapsto L^2$ is defined as

$$\mathcal{K} f(s) = \int_M K(s,t) f(t) \mu(\mathrm{d}t),$$

corresponding to the covariance kernel *K* of *Z* [for properties of covariance kernels, see Shorack and Wellner (1986), p.207]. The task of finding eigenvalues and eigenfunctions is often referred to as the kernel eigenproblem; see Williams and Shawe-Taylor (2003). Solving (1) analytically is generally considered challenging. For a list of solutions for specific choices of $\mathcal{K}$, *M*, and μ, see Appendix A in Fasshauer and McCourt (2016). From now on, we assume that μ is defined by a non-negative weight function *w* on *M*, such that $\mu(\mathrm{d}t) = w(t)\mathrm{d}t$ and $\int_M t w(t) \mathrm{d}t < \infty$. We will also consider an example where $M = \mathbb{N}_0 = \{0, 1, 2, \ldots\}$ and μ has a density (in the Radon-Nikodym sense) with respect to the counting measure.

3 The Rayleigh–Ritz method

As mentioned in the Introduction, let $K(s,t)$ be a covariance kernel. Given that it is symmetric and nonnegative definite, its eigenvalues are real and nonnegative. Throughout this paper, we assume implicitly that $\|K\|^2_{\mathcal{H}} = \int_M \int_M K(s,t)^2 \mu(\mathrm{d}s)\mu(\mathrm{d}t) < \infty$, which in turn implies that $\int_M K(t,t)\mu(\mathrm{d}t) < \infty$.

For the sake of completeness, this section reformulates the Rayleigh–Ritz method. Let λ_1 denote the largest eigenvalue of the covariance kernel K and let f_1 be the associated normalized eigenfunction, such that $\|f_1\|^2_{L^2} = 1$. Let $\{\psi_j\}_{j\geq 1}$ represent an orthonormal basis of L^2. The core idea is to approximate f_1 by $f_{1n} = \sum_{j=1}^n \alpha_j \psi_j$, with $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$, for some $n \in \mathbb{N}$. In this approximation, $\alpha_j = \langle f_1, \psi_j \rangle$, for $1 \leq j \leq n$, which are unknown quantities (as f_1 is unknown). To estimate $\lambda_1 = \langle \mathcal{K} f_1, f_1 \rangle$, we maximize the function

$$\langle \mathcal{K} f_{1n}, f_{1n} \rangle = \left\langle K \sum_{j=1}^n \alpha_j \psi_j, \sum_{j=1}^n \alpha_j \psi_j \right\rangle = \sum_{j,k=1}^n K_{jk} \alpha_j \alpha_k,$$

subject to

$$\left\| \sum_{j=1}^n \alpha_j \psi_j \right\|^2_{L^2} = \sum_{j=1}^n \alpha_j^2 = 1,$$

where $K_{jk} = \langle \mathcal{K}\psi_j, \psi_k \rangle = \langle \psi_j, \mathcal{K}\psi_k \rangle$. Using the method of Lagrangian multipliers, we set

$$\mathrm{L}(x) = \sum_{j,k=1}^n K_{jk}\alpha_j\alpha_k - x\delta_{jk}\alpha_j\alpha_k, \quad x \in \mathbb{R},$$

where δ_{jk} is the Kronecker delta. The extremal values of $\alpha_1, \ldots, \alpha_k$ are determined from the equations $\partial \mathrm{L}(x)/\partial \alpha_j = 0$ for $j = 1, \ldots, n$, leading to

$$\sum_{k=1}^n K_{jk}\alpha_k - x\alpha_j = 0, \quad j = 1, \ldots, n.$$

This system of equations has a nontrivial solution if and only if the determinant

$$\det \begin{pmatrix} K_{11} - x & K_{12} & \cdots & K_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ K_{n1} & K_{n2} & \cdots & K_{nn} - x \end{pmatrix} = 0. \tag{2}$$

Thus, the method calculates the eigenvalues of the $n \times n$ matrix $M_n = (K_{jk})_{1 \leq j,k \leq n}$. For $1 \leq m \leq n$, the *m* largest eigenvalues of M_n, denoted by $\widehat{\lambda}_1, \ldots, \widehat{\lambda}_m$, estimate the *m* largest eigenvalues of the covariance kernel *K*, denoted by $\lambda_1, \ldots, \lambda_m$. Larger values of *n* will improve the approximation. In fact, Theorem 3.1 below shows that $|\widehat{\lambda}_i - \lambda_i| \to 0$ as $n \to \infty$. Before stating and proving this result, we first see that the Rayleigh–Ritz method approximates not only f_1 by f_{1n}, but also the kernel *K*.

If $\{\psi_j(s),\, s \in M\}_{j \geq 1}$ is an orthonormal basis of L^2, then $\{\psi_j(s)\psi_k(t),\, s, t \in M\}_{j,k \geq 1}$ forms an orthonormal set in the (separable) Hilbert space $\mathcal{H} = \{h : M \times M \to \mathbb{R} \,|\, \|h\|_{\mathcal{H}}^2 = \int_M \int_M h(s,t)^2 \mu(\mathrm{d}s)\mu(\mathrm{d}t) < \infty\}$, with the inner product $\langle h_1, h_2 \rangle_{\mathcal{H}} = \int_M \int_M h_1(s,t) h_2(s,t) \mu(\mathrm{d}s)\mu(\mathrm{d}t)$, for $h_1, h_2 \in \mathcal{H}$. Thus, one can approximate the covariance kernel *K*(*s*, *t*) by

$$K_n(s,t) = \sum_{j,k=1}^{n} K_{jk} \psi_j(s) \psi_k(t). \tag{3}$$

It is easy to see that K_n is symmetric and nonnegative definite, hence its eigenvalues are real and nonnegative. Moreover, it has at most *n* positive eigenvalues. Let $\mathcal{K}_n$ denote the operator analogous to $\mathcal{K}$, defined using K_n instead of *K*. It is straightforward to verify that

$$\langle \mathcal{K} f_{1n}, f_{1n} \rangle = \langle \mathcal{K}_n f_1, f_1 \rangle = \langle \mathcal{K}_n f_{1n}, f_{1n} \rangle.$$

Therefore, the Rayleigh–Ritz method essentially approximates both f_1 by f_{1n} and *K* by K_n.

The next result shows that the solutions of Eq. (2) provide good approximations to the eigenvalues of *K*.

Theorem 3.1 *Let $\{\lambda_i\}_{i \geq 1}$ denote the eigenvalues of K, arranged in decreasing order and repeated according to their multiplicity. Let $\{\widehat{\lambda}_i\}_{i=1}^{n}$ represent the solutions (in x) of Eq. (2), also arranged in decreasing order and repeated according to their multiplicity. Then, $|\widehat{\lambda}_i - \lambda_i| \to 0$, as $n \to \infty$, for each i.*

Proof Section 7.7 of Kanwal (1971), shows that

$$\|K - K_n\|_{\mathcal{H}} \to 0, \qquad \text{as } n \to \infty, \tag{4}$$

where K_n is as defined in (3). Lemma 2.2 in Horváth and Kokoszka (2012) states that

$$|\widehat{\lambda}_i - \lambda_i| \leq \|\mathcal{K} - \mathcal{K}_n\|_{\mathcal{L}}, \tag{5}$$

where $\|\Psi\|_{\mathcal{L}} = \sup\{\|\Psi(x)\|_{L^2} : \|x\|_{L^2} = 1\}$. We also have that [see e.g. (2.3) and (2.5) in Horváth and Kokoszka (2012)] that

$$\|\mathcal{K} - \mathcal{K}_n\|_{\mathcal{L}} \leq \|K - K_n\|_{\mathcal{H}}. \tag{6}$$

The result follows from (4)–(6). □

Remark 3.2 It is desirable to have a bound for $|\widehat{\lambda}_i - \lambda_i|$. From the proof of Theorem 3.1, we have $|\widehat{\lambda}_i - \lambda_i| \leq \|\mathcal{K} - \mathcal{K}_n\|_{\mathcal{L}}$, for each *i*, but $\|\mathcal{K} - \mathcal{K}_n\|_{\mathcal{L}}$ cannot be easily calculated. From (5), we also have

$$|\widehat{\lambda}_i - \lambda_i| \leq \|K - K_n\|_{\mathcal{H}} = \left(\|K\|^2_{\mathcal{H}} - \sum_{j,k=1}^{n} K^2_{ij} \right)^{1/2}, \tag{7}$$

The right-hand side of the above inequality can be readily calculated but, as it will be seen in Example 4.1 of the next section, the bound is not sharp.

Let $E = (e_{ij})$ the $n \times n$ matrix whose columns contain the eigenvectors of M_n, which are assumed to be normalized so that all of them have Euclidean norm equal to 1 and arranged so that $M_n(e_{i1}, \ldots, e_{in})^\top = \widehat{\lambda}_i(e_{i1}, \ldots, e_{in})^\top$ and $\widehat{\lambda}_1 \geq \ldots \geq \widehat{\lambda}_n$. Let $f_{in}(x) = \sum_{j=1}^{n} e_{ij}\psi_j(x), 1 \leq i \leq n$. As observed before for $i = 1$, the Rayleigh–Ritz method approximates f_i by f_{in}. One may wonder if f_i and f_{in} are close. In the following we examine this point.

In order to approximate the eigenfunctions, they must be identifiable, so we assume that

$$\lambda_1 > \lambda_2 > \ldots > \lambda_p, \tag{8}$$

for some fixed $p \in \mathbb{N}$, where $\{\lambda_i\}_{i \geq 1}$ denote the eigenvalues of *K*, arranged in decreasing order and repeated according to their multiplicity. Notice that if *e* is an eigenvector of M_n then so is $-e$. Because of this reason, we define

$$c_{in} = \operatorname{sign}(\langle f_i, f_{in} \rangle), \quad f'_{in} = c_{in} f_{in}.$$

The next result shows that if (8) holds, then $f'_{1n}, \ldots, f'_{pn}$ provide good approximations to $f_1, \ldots, f_p$, respectively. Notice that, in practice, c_{in} is an unknown quantity. Therefore, $f_{1n}, \ldots, f_{pn}$ are good subrogates of $f_1, \ldots, f_p$, except for the sign.

Theorem 3.3 *Suppose that (8) holds. Then,* $|f'_{in} - f_i| \to 0$*, as* $n \to \infty$*,* $1 \leq i \leq p$*.*

Proof If (8) holds, Lemma 2.3 of Horváth and Kokoszka (2012) states that

$$\|f'_{in} - f_i\| \leq \frac{2\sqrt{2}}{\alpha_j} \|\mathcal{K} - \mathcal{K}_n\|_{\mathcal{L}}, \tag{9}$$

where $\|\Psi\|_{\mathcal{L}}$ is as defined in the proof of Theorem 3.1, $\alpha_1 = \lambda_2 - \lambda_1$ and $\alpha_j = \min(\lambda_{j-1} - \lambda_j, \lambda_j - \lambda_{j+1})$, $2 \le j \le p$. The result follows from the previous inequality, (4), (6) and (8). □

Remark 3.4 The proof of Theorem 3.3 provides a bound for $\|f'_{in} - f_i\|$ given in (9). The term $\|\mathcal{K} - \mathcal{K}_n\|_{\mathcal{L}}$ on the right-hand side of (9) can be dealt with as in Remark 3.2. Nevertheless, the term $2\sqrt{2}/\alpha_j$ is unknown.

It is evident that the choice of the orthonormal set $\{\psi_j\}_{j\ge1}$ depends on the support *M* and the weight function *w*(*t*). We provide examples for the most common supports and weight functions, utilizing orthogonal polynomials. For a list of classical polynomials, see Table 1.1 on p. 29 of Gautschi (2004).

4 Applications

In this section, we employ the Rayleigh–Ritz method to address both resolved and open eigenvalue problems available in the literature. The solved cases serve to illustrate the accuracy of the approximations obtained, while the unsolved problems demonstrate the versatility of the method. We provide appropriate orthonormal bases tailored to the different supports and weight functions of the Hilbert spaces L^2. Throughout, we use the notation $u \wedge v = \min(u, v)$ and $u \vee v = \max(u, v)$ for $u, v \in \mathbb{R}$.

4.1 Support $M = [0, 1]$

Let the integration domain be $M = [0, 1]$. We focus on the weight function $w(t) = 1$. Accordingly, a suitable set of orthonormal polynomials can be defined as follows:

$$\phi_k(x) = \sqrt{2k+1}\, P_k(2x - 1), \quad x \in [0, 1],$$

where

$$P_k(x) = \sum_{j=0}^{\lfloor k/2 \rfloor} (-1)^j \frac{(2k-2j)!}{(k-j)!(k-2j)!\, j!\, 2^k} x^{k-2j}$$

denotes the Legendre polynomial of degree k. In what follows, $n = n_0$ means that we have taken all polynomials with degree less than or equal to n_0, that in the present case implies that we are considering the $n_0 + 1$ elements $\phi_0, \phi_1, \ldots, \phi_{n_0}$.

Example 4.1 (Non-parametric goodness-of-fit tests)

- The traditional Cramér-von Mises test statistic possesses the covariance kernel

$$K(s,t) = s \wedge t - st, \quad s,t \in [0,1], \tag{10}$$

refer to Shorack and Wellner (1986, p. 14). It is a well-established fact that the sequence of eigenvalues is given by $\lambda_j = 1/j^2\pi^2$, $j = 1, 2, \ldots$, [see e.g. Proposition 1 on pag. 213–214 of Shorack and Wellner (1986)]. Table 1 presents the estimators of the five largest eigenvalues of kernel (10), for $3 \le n \le 15$, along with the actual values. Throughout this paper, e-*j* stands for 10^{-j}. Observing this table, we notice that, with the precision in it, for $n \ge 4$ the estimator of the largest eigenvalue matches the actual value; $n \ge 9$ for the second; $n \ge 10$ for the third, and so forth. This table also displays the value on the right-hand side of (7) as well as $\max_i |\widehat{\lambda}_i - \lambda_i|$. Notice that the bound in (7) is, in most cases, larger than 2.5 times the true error.

An anonymous referee suggested that we investigate the behavior of the method when *m* is close to *n*, as well as its performance when both *n* and *m* are large. To this end, Table 2 reports the values of $\widehat{\lambda}_i$ for $6 \le i \le 15$, and $5 \le n \le 22$. By looking at Tables 1 and 2 (see also Fig. 2), we observe that: $\hat{\lambda}_m$ is not a good estimator of λ_m, when *m* and *n* are close; the estimators for the largest eigenvalues exhibit rapid convergence, whereas convergence is slower for the smaller eigenvalues. These patterns are not unique to this particular example; they have been consistently observed in other cases as well. For brevity, we only present results for the current setting. Furthermore, increasing *n* excessively may degrade estimation accuracy due to error accumulation: recall that the quantities $\{K_{jk},\, 1 \le j,k \le n\}$ are integrals, some of which can be evaluated analytically, while others require numerical approximation. Since we are estimating relatively small parameters, the numerical integration error can occasionally exceed the magnitude of the quantity being estimated.

The eigenfunctions of the kernel (10) are well known [see, for instance, Proposition 1 on pp. 213–214 of Shorack and Wellner (1986)], and are given by $f_j(t) = \sqrt{2}\sin(j\pi t)$, for $0 \le t \le 1$ and $j = 1, 2, \ldots$. Table 3 reports the values of $\|f'_{in} - f_i\|$ for $1 \le i \le 5$ and $3 \le n \le 10$. As in the case of the eigenvalues, the table clearly indicates a rapid convergence of the eigenfunctions corresponding to the largest eigenvalues. This behavior is further illustrated in Fig. 1, which shows the functions f_2, f'_{2n} for $n = 3, 5$, f_3, f'_{3n}, and f_4, f'_{4n} for $n = 3, 5, 7, 9$.

- In Henze and Nikitin (2000), Equation (3.2), the covariance kernel is characterized by

$$\overline{K}_0(s,t) = \frac{st(s \wedge t)}{2} - \frac{(s \wedge t)^3}{6} - \frac{s^2t^2}{4}, \quad s,t \in [0,1]. \tag{11}$$

Table 4 presents the estimators of the five largest eigenvalues of kernel (11), for $4 \le n \le 16$, along with the actual values, which are taken from Table 3.1 of Henze and Nikitin (2000). Observing this table, we notice that, with the precision in it, for $n \ge 5$ the estimator of the largest eigenvalue matches the actual value. As observed before, larger eigenvalues require larger n, with exception of the third eigenvalue marked by*. In this case we conjecture that the authors of Henze and Nikitin (2000) made a numerical or typographical error in the fourth decimal place.

Table 1 Estimators of the five largest eigenvalues of kernel (10) obtained for $3 \leq n \leq 15$ with the Rayleigh–Ritz method. The column headed $\|K - K_n\|_{\mathcal{H}}$ displays the value of the right-hand side of (7)

n	$\widehat{\lambda}_1$	$\widehat{\lambda}_2$	$\widehat{\lambda}_3$	$\widehat{\lambda}_4$	$\widehat{\lambda}_5$	$\|K - K_n\|_{\mathcal{H}}$	$\max_i \|\hat{\lambda}_i - \lambda_i\|$
3	1.012648e–1	2.514783e–2	0.587803e–2	2.629952e–3	0.000000e–3	13.515386e–3	5.379875e–3
4	1.013212e–1	2.514783e–2	1.096185e–2	2.629952e–3	1.353354e–3	9.154818e–3	3.702622e–3
5	1.013212e–1	2.532945e–2	1.096185e–2	5.955316e–3	1.353354e–3	6.745825e–3	2.699493e–3
6	1.013212e–1	2.532945e–2	1.125318e–2	5.955316e–3	3.626214e–3	5.243460e–3	2.047961e–3
7	1.013212e–1	2.533029e–2	1.125318e–2	6.318910e–3	3.626214e–3	4.229955e–3	1.601687e–3
8	1.013212e–1	2.533029e–2	1.125789e–2	6.318910e–3	4.025014e–3	3.507253e–3	1.283693e–3
9	1.013212e–1	2.533030e–2	1.125789e–2	6.332401e–3	4.025014e–3	2.970122e–3	1.049834e–3
10	1.013212e–1	2.533030e–2	1.125791e–2	6.332401e–3	4.052147e–3	2.557849e–3	0.873283e–3
11	1.013212e–1	2.533030e–2	1.125791e–2	6.332573e–3	4.052147e–3	2.233147e–3	0.737003e–3
12	1.013212e–1	2.533030e–2	1.125791e–2	6.332573e–3	4.052840e–3	1.971950e–3	0.629780e–3
13	1.013212e–1	2.533030e–2	1.125791e–2	6.332574e–3	4.052840e–3	1.758096e–3	0.544009e–3
14	1.013212e–1	2.533030e–2	1.125791e–2	6.332574e–3	4.052847e–3	1.580361e–3	0.474394e–3
15	1.013212e–1	2.533030e–2	1.125791e–2	6.332574e–3	4.052847e–3	1.430730e–3	0.417163e–3
True	1.013212e–1	2.533030e–2	1.125791e–2	6.332574e–3	4.052847e–3		

- In Ebner et al. (2022), Theorem 2.1, the covariance kernel is defined as

$$K_Z(s,t) = \frac{1-(2(s \vee t)-1)^3}{6} - st(1-s)(1-t), \quad s,t \in (0,1). \tag{12}$$

Table 5 presents the estimators of the five largest eigenvalues of kernel (12), for $3 \leq n \leq 15$. Similar to the previous two cases, we notice a quick convergence of the first few eigenvalues.–

4.2 Support $M = [0, \infty)$

If the domain of integration is $M = [0, \infty)$, we concentrate on the weight function $w_\gamma(t) = e^{-\gamma t}$ for a positive parameter γ. A suitable set of orthonormal polynomials can then be expressed as

$$\phi_k(x;\gamma) = \sqrt{\gamma} L_k(\gamma x), \quad x \in [0, \infty),$$

where $L_k(x) = \sum_{j=0}^{k} \left(\begin{array}{c} k \\ j \end{array} \right) \frac{(-1)^j}{j!} x^j$ represents the Laguerre polynomial of degree k.

Example 4.2 (Testing Exponentiality)

- In Baringhaus and Henze (2008), the covariance kernel is specified in (7) as

$$\rho(s,t) = \min(1-e^{-s}, 1-e^{-t}) - (1-e^{-s})(1-e^{-t}), \quad s,t \geq 0. \tag{13}$$

In this scenario, the formulas for the eigenvalues are explicitly provided, see Theorem 2 in Baringhaus and Henze (2008). It's worth noting that in Ebner (2023), the author demonstrates that these eigenvalues also correspond to the covariance kernel

$$K_0(s,t) = e^{-s \vee t} - e^{-(s+t)}, \quad s,t \geq 0, \tag{14}$$

and the first twenty eigenvalues are tabulated in Table 2 in Ebner (2023). Table 6 presents the estimators of the two largest eigenvalues of kernel (14), for $n = 10(5)30$, and $\gamma = 1, 2$, along with the actual values, which are taken from Table 2 in Ebner (2023).

- In Klar (2001), Theorem 2.1, the covariance kernel is defined as

$$K(s,t) = (|s-t|+2)e^{-s \vee t} - (s+t+st+2)e^{-(s+t)}, \quad s,t \geq 0. \tag{15}$$

Table 7 presents the estimators of the two largest eigenvalues of kernel (15), for $n = 10(5)30$, and $\gamma = 0.5, 1, 1.5$.

Table 2 Estimators of $\lambda_6 - \lambda_{15}$ of kernel (10) obtained for $5 \leq n \leq 22$ with the Rayleigh–Ritz method

n	$\widehat{\lambda}_6$	$\widehat{\lambda}_7$	$\widehat{\lambda}_8$	$\widehat{\lambda}_9$	$\widehat{\lambda}_{10}$	$\widehat{\lambda}_{11}$	$\widehat{\lambda}_{12}$	$\widehat{\lambda}_{13}$	$\widehat{\lambda}_{14}$	$\widehat{\lambda}_{15}$
5	0.766517e−3	0.000000e−3	0.000000e−3	0.000000e−3	0.000000e−3	0.000000e−4	0.000000e−4	0.000000e−4	0.000000e−4	0.000000e–4
6	0.766517e−3	0.466093e−3	0.000000e−3	0.000000e−3	0.000000e−3	0.000000e−4	0.000000e−4	0.000000e−4	0.000000e−4	0.000000e–4
7	2.365070e−3	0.466093e−3	0.299451e−3	0.000000e−3	0.000000e−3	0.000000e−4	0.000000e−4	0.000000e−4	0.000000e−4	0.000000e–4
8	2.365070e−3	1.615925e−3	0.299451e−3	0.201045e−3	0.000000e−3	0.000000e−4	0.000000e−4	0.000000e−4	0.000000e−4	0.000000e–4
9	2.768716e−3	1.615925e−3	1.142944e−3	0.201045e−3	0.139929e−3	0.000000e−4	0.000000e−4	0.000000e−4	0.000000e−4	0.000000e–4
10	2.768716e−3	2.002557e−3	1.142944e−3	0.831058e−3	0.139929e−3	1.003626e−4	0.000000e−4	0.000000e−4	0.000000e−4	0.000000e–4
11	2.812482e−3	2.002557e−3	1.499078e−3	0.831058e−3	0.618398e−3	1.003626e−4	0.738395e−4	0.000000e−4	0.000000e−4	0.000000e–4
12	2.812482e−3	2.063326e−3	1.499078e−3	1.150215e−3	0.618398e−3	4.693716e−4	0.738395e−4	0.555247e−4	0.000000e−4	0.000000e–4
13	2.814439e−3	2.063326e−3	1.574828e−3	1.150215e−3	0.899155e−3	4.693716e−4	3.624786e−4	0.555247e−4	0.425507e−4	0.000000e–4
14	2.814439e−3	2.067641e−3	1.574828e−3	1.237322e−3	0.899155e−3	7.134470e−4	3.624786e−4	2.842389e−4	0.425507e−5	0.331533e–4
15	2.814477e−3	2.067641e−3	1.582752e−3	1.237322e−3	0.993319e−3	7.134470e−4	5.732227e−4	2.842389e−4	2.259397e−4	0.331533e–4
16	2.814477e−3	2.067777e−3	1.582752e−3	1.249957e−3	0.993319e−3	8.104911e−4	5.732227e−4	4.655949e−4	2.259397e−4	1.817913e–4
17	2.814477e−3	2.067777e−3	1.583134e−3	1.249957e−3	1.011353e−3	8.104911e−4	6.696610e−4	4.655949e−4	3.818958e−4	1.817913e–4
18	2.814477e−3	2.067779e−3	1.583134e−3	1.250847e−3	1.011353e−3	8.340574e−4	6.696610e−4	5.589522e−4	3.818958e−4	3.162599e–4
19	2.814477e−3	2.067779e−3	1.583143e−3	1.250847e−3	1.013102e−3	8.340576e−4	6.977392e−4	5.589535e−4	4.684666e−4	3.162608e–4
20	2.814477e−3	2.067779e−3	1.583143e−3	1.250876e−3	1.013103e−3	8.369528e−4	6.977529e−4	5.895202e−4	4.685014e−4	3.936315e–4
21	2.814477e−3	2.067779e−3	1.583143e−3	1.250876e−3	1.013133e−3	8.369780e−4	6.992564e−4	5.898175e−4	4.783055e−4	3.941296e–4
22	2.814477e−3	2.067779e−3	1.583144e−3	1.250876e−3	1.013142e−3	8.369658e−4	6.994945e−4	5.900715e−4	4.798897e−4	3.973553e–4
True	2.814477e−3	2.067779e−3	1.583144e−3	1.250879e−3	1.013212e−3	8.373652e−4	7.036193e−4	5.995336e−4	5.169448e−4	4.503164e–4

4.3 Support $M = \mathbb{R}$

If the domain of integration is denoted by $M = \mathbb{R}$, we concentrate on the weight function $w_\gamma(t) = e^{-\gamma t^2}$ for a positive parameter $\gamma > 0$. A suitable set of orthonormal polynomials can then be expressed as

$$\phi_k(x;\gamma) = \left(2^k\, k!\sqrt{\pi/\gamma}\right)^{-1/2} H_k\left(\sqrt{\gamma}x\right), \quad x \in \mathbb{R}, \tag{16}$$

where $H_k(x) = (-1)^k \exp(x^2)\frac{\mathrm{d}^k}{\mathrm{d}\,x^k}\exp(-x^2)$ represents the Hermite polynomial of degree *k*.

Example 4.3 (Normality Test)

- In Ebner (2020), Theorem 2.2, the covariance kernel is defined as

$$K_Z(s,t) = (st+1)\exp\left(-\frac{(s-t)^2}{2}\right) - (2st+1)\exp\left(-\frac{s^2+t^2}{2}\right), \quad s,t \in \mathbb{R}. \tag{17}$$

The exact values of the four cumulants of the distribution of $T(\gamma) = \|Z\|^2_{L^2}$ with $w = \varphi_\gamma$ (denoted as $\kappa_1(\gamma) = \mathbb{E}(T(\gamma))$, $\kappa_2(\gamma) = \mathbb{E}(T(\gamma) - \kappa_1(\gamma))^2$, $\kappa_3(\gamma) = \mathbb{E}(T(\gamma) - \kappa_1(\gamma))^3$ and $\kappa_4(\gamma) = \mathbb{E}(T(\gamma) - \kappa_1(\gamma))^4 - 3\mathbb{E}^2(T(\gamma) - \kappa_1(\gamma))^2)$ have been provided in Ebner (2020) for several values of γ. As seen in Sect. 6.1,

$$\begin{aligned} \kappa_1(\gamma) &= \sum_{j=1}^{\infty}\lambda_j(\gamma), & \kappa_2(\gamma) &= 2\sum_{j=1}^{\infty}\lambda_j^2(\gamma), \\ \kappa_3(\gamma) &= 8\sum_{j=1}^{\infty}\lambda_j^3(\gamma), & \kappa_4(\gamma) &= 48\sum_{j=1}^{\infty}\lambda_j^4(\gamma), \end{aligned}$$

where the set $\{\lambda_j(\gamma)\}_{j\ge 1}$ denotes the solutions in λ of Eq. (1) with $\mu(\mathrm{d}t) = w_\gamma(t)\mathrm{d}t$. To verify the accuracy of the solutions of Eq. (2), we calculated the approximations

$$\begin{aligned} \widehat{\kappa}_{n1}(\gamma) &= \sum_{j=1}^{m}\widehat{\lambda}_j, & \widehat{\kappa}_{n2}(\gamma) &= 2\sum_{j=1}^{m}\widehat{\lambda}_j^2, \\ \widehat{\kappa}_{n3}(\gamma) &= 8\sum_{j=1}^{m}\widehat{\lambda}_j^3, & \widehat{\kappa}_{n4}(\gamma) &= 48\sum_{j=1}^{m}\widehat{\lambda}_j^4, \end{aligned} \tag{18}$$

where *m* is the number of non-null solutions of Eq. (2) obtained by considering the first *n* elements of the basis (16). Table 8 presents $\kappa_i(\gamma)$ and $\widehat{\kappa}_{ni}(\gamma)$ for $\gamma = 0.5, 1, 2,$ and $m = n = 10(5)30$.

- In Henze et al. (2019), Theorem 5.1, the covariance kernel is characterized as

Table 3 Values of $\|f'_{in} - f_i\|$ for the kernel (11), $1 \le i \le 5$ and $3 \le n \le 10$

n	$\|f'_{1n} - f_1\|$	$\|f'_{2n} - f_2\|$	$\|f'_{3n} - f_3\|$	$\|f'_{4n} - f_4\|$	$\|f'_{5n} - f_5\|$
3	2.444705e−2	9.437187e−2	8.329145e−1	9.687312e−1	
4	5.220099e−4	9.437187e−2	1.943579e−1	9.687312e−1	1.078522
5	5.220099e−4	6.096467e−3	1.943579e−1	3.132346e−1	1.078522
6	5.876814e−6	6.096468e−3	2.279650e−2	3.132346e−1	4.417897e–1
7	5.876814e−6	2.215575e−4	2.279650e−2	5.511943e−2	4.417897e–1
8	4.085192e−8	2.215575e−4	1.561379e−3	5.511943e−2	1.055505e–1
9	4.085192e−8	5.174104e−6	1.561379e−3	5.779082e−3	1.055505e–1
10	1.928001e−10	5.174104e−6	7.078110e−5	5.779082e−3	1.525461e–2

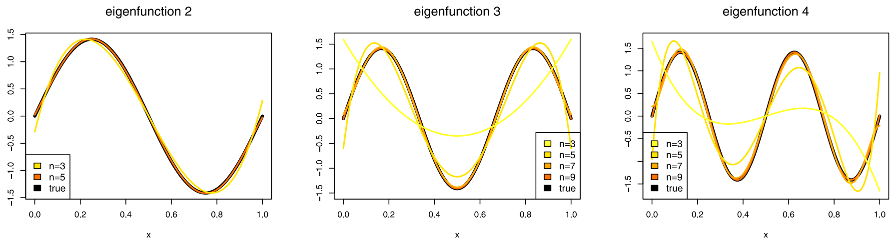

Fig. 1 $f_2, f'_{2n}, n = 3, 5, f_3, f'_{3n}$, and $f_4, f'_{4n}, n = 3, 5, 7, 9$, for kernel (10)

$$C(s,t) = \exp(st) + \frac{1}{2}(\exp(st) + \exp(-st)) + 2\cos(st) - st - 4, \quad s,t \in \mathbb{R}. \tag{19}$$

Note that the authors in this case assume $\gamma > 1$. Table 9 presents the estimators of the two largest eigenvalues of kernel (19), for $5 \le n \le 13$, and $\gamma = 1.5, 2, 3$. In this case, the convergence is notably quick, especially for $\gamma = 3$.

- In Dörr et al. (2021), Theorem 5, the covariance kernel is characterized as

$$\begin{aligned} K(s,t) = &\exp\left(-\frac{(s-t)^2}{2}\right)\left(\left((s-t)^2-3\right)^2-6\right) \\ &+\exp\left(-\frac{s^2+t^2}{2}\right)\left(-\frac{s^2t^2\left(s^2-5\right)\left(t^2-5\right)}{2}+6(s^2+t^2)-s^4-t^4\right. \\ &\left.-s^2t^2-st\left(s^2-3\right)\left(t^2-3\right)-3\right), \end{aligned} \tag{20}$$

- $s,t \in \mathbb{R}$. For $n = 15$, the largest eigenvalue $\widehat{\lambda}_1(\gamma)$ is approximated by $\widehat{\lambda}_1(1/2) = 0.966103408152626$, $\widehat{\lambda}_1(1) = 0.600668091541773$, $\widehat{\lambda}_1(2) = 0.398015644894253$, and $\widehat{\lambda}_1(3) = 0.310469662783734$.

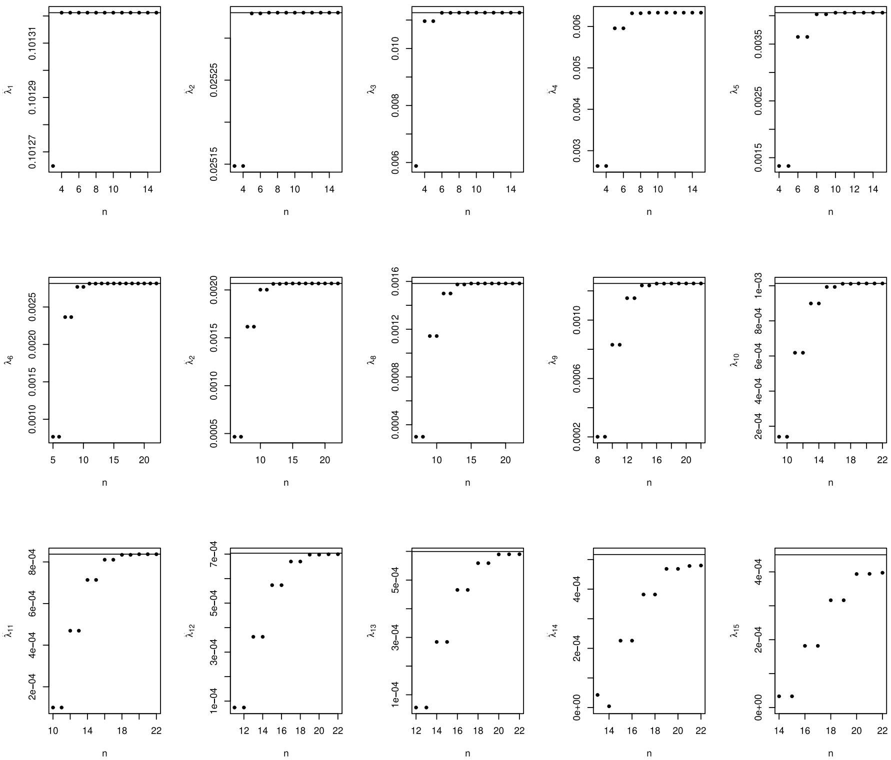

Fig. 2 Plot of the eigenvalues presented in Tables 1 and 2: *horizontal lines* are true values

Table 4 Estimators of the five largest eigenvalues of kernel (11) obtained for $3 \leq n \leq 15$ with the Rayleigh–Ritz method

n	$\widehat{\lambda}_1$	$\widehat{\lambda}_2$	$\widehat{\lambda}_3$	$\widehat{\lambda}_4$	$\widehat{\lambda}_5$
3	3.196304e–2	1.058647e–3	1.564775e–4	0.000000e–5	0.000000e–5
4	3.196394e–2	1.093812e–3	1.633718e–4	3.999208e–5	0.000000e–5
5	3.196395e–2	1.094444e–3	1.787795e–4	4.309337e–5	1.330054e–5
6	3.196395e–2	1.094568e–3	1.792433e–4	5.121087e–5	1.484606e–5
7	3.196395e–2	1.094568e–3	1.794972e–4	5.155691e–5	1.950214e–5
8	3.196395e–2	1.094569e–3	1.795006e–4	5.189992e–5	1.976122e–5
9	3.196395e–2	1.094569e–3	1.795022e–4	5.190770e–5	2.014210e–5
10	3.196395e–2	1.094569e–3	1.795022e–4	5.191281e–5	2.015456e–5
11	3.196395e–2	1.094569e–3	1.795022e–4	5.191289e–5	2.016592e–5
12	3.196395e–2	1.094569e–3	1.795022e–4	5.191292e–5	2.016616e–5
13	3.196395e–2	1.094569e–3	1.795022e–4	5.191292e–5	2.016629e–5
14	3.196395e–2	1.094569e–3	1.795022e–4	5.191292e–5	2.016629e–5
15	3.196395e–2	1.094569e–3	1.795022e–4	5.191292e–5	2.016629e–5
true	3.196395e–2	1.094569e–3	1.795105e–4*	5.191292e–5	2.016629e–5

Table 5 Estimators of the five largest eigenvalues of kernel (12) obtained for $3 \le n \le 15$ with the Rayleigh–Ritz method

n	$\widehat{\lambda}_1$	$\widehat{\lambda}_2$	$\widehat{\lambda}_3$	$\widehat{\lambda}_4$	$\widehat{\lambda}_5$
3	1.155505e–1	6.173688e–3	2.232223e–3	0.000000e–3	0.0000000e–3
4	1.157325e–1	6.553968e–3	2.287779e–3	1.615387e–3	0.0000000e–3
5	1.157342e–1	6.936156e–3	2.451756e–3	1.636573e–3	0.8705469e–3
6	1.157342e–1	6.941384e–3	2.699774e–3	1.851195e–3	0.8713323e–3
7	1.157342e–1	6.941476e–3	2.721395e–3	1.946449e–3	1.0884916e–3
8	1.157342e–1	6.942877e–3	2.723814e–3	1.971728e–3	1.1784099e–3
9	1.157342e–1	6.943677e–3	2.725788e–3	1.973082e–3	1.1888306e–3
10	1.157342e–1	6.943684e–3	2.728232e–3	1.976252e–3	1.1888330e–3
11	1.157342e–1	6.943684e–3	2.728373e–3	1.977335e–3	1.1955965e–3
12	1.157342e–1	6.943684e–3	2.728383e–3	1.977537e–3	1.1976705e–3
13	1.157342e–1	6.943685e–3	2.728387e–3	1.977544e–3	1.1978559e–3
14	1.157342e–1	6.943685e–3	2.728391e–3	1.977555e–3	1.1978559e–3
15	1.157342e–1	6.943685e–3	2.728392e–3	1.977558e–3	1.1979070e–3

Table 6 Estimators of the two largest eigenvalues of kernel (14) obtained for $n = 10(5)30$ and $\gamma = 1, 2$ with the Rayleigh–Ritz method

	$\gamma = 1$		$\gamma = 2$	
n	$\widehat{\lambda}_1$	$\widehat{\lambda}_2$	$\widehat{\lambda}_1$	$\widehat{\lambda}_2$
10	1.012749e−1	2.346677e−2	5.274782e−2	1.210710e–2
15	1.013091e−1	2.517918e−2	5.275176e−2	1.218831e–2
20	1.013168e−1	2.527533e−2	5.275292e−2	1.220204e–2
25	1.013204e−1	2.529163e−2	5.275300e−2	1.221051e–2
30	1.013211e−1	2.531704e−2	5.275301e−2	1.221147e–2
true	1.013212e−1	2.533030e−2	5.275301e−2	1.221201e–2

Table 7 Estimators of the two largest eigenvalues of kernel (15) obtained for $n = 10(5)30$ and $\gamma = 0.5, 1, 1.5$ with the Rayleigh–Ritz method

	$\gamma = 0.5$		$\gamma = 1$		$\gamma = 1.5$	
n	$\widehat{\lambda}_1$	$\widehat{\lambda}_2$	$\widehat{\lambda}_1$	$\widehat{\lambda}_2$	$\widehat{\lambda}_1$	$\widehat{\lambda}_2$
10	6.905560e−2	0.9839087e−2	3.091694e−2	4.061322e−3	1.581485e−2	1.958083e−3
15	6.949381e−2	1.0149090e−2	3.094987e−2	4.154889e−3	1.581630e−2	1.980193e−3
20	6.963910e−2	1.0301102e−2	3.095009e−2	4.168487e−3	1.581631e−2	1.980447e−3
25	6.965260e−2	1.0420650e−2	3.095011e−2	4.168621e−3	1.581631e−2	1.980510e−3
30	6.965294e−2	1.0445512e−2	3.095012e−2	4.168761e−3	1.581631e−2	1.980516e−3

4.4 Support $M = \mathbb{N}_0$

If the domain of integration is $M = \mathbb{N}_0$, we concentrate on the measure with derivative with respect to the counting measure given by $w_\varrho(t) = e^{-\varrho}\varrho^t/t!$ for a positive parameter $\varrho > 0$. In other words, we consider $\ell^2(\varrho)$, the (separable) Hilbert space of all infinite sequences $a = (a_0, a_1, \ldots)$ of complex numbers such that $\sum_{t\ge 0} |a_t|^2 w_\varrho(t) < \infty$, with the inner product defined as

Table 8 First four cumulants and their Rayleigh–Ritz estimators of the distribution of $T(\gamma) = \|Z\|^2_{L^2}$, Z a centered Gaussian process with covariance kernel K_Z in (17)

γ		n	$i = 1$	$i = 2$	$i = 3$	$i = 4$
0.5	$\kappa_i(\gamma)$		2.60125	4.71530	20.04364	133.19802
	$\widehat{\kappa}_{ni}(\gamma)$	10	2.57020	4.69630	19.97998	132.77912
		15	2.59684	4.71475	20.04278	133.19340
		20	2.60065	4.71527	20.04361	133.19786
		25	2.60117	4.71530	20.04364	133.19807
		30	2.60124	4.71530	20.04365	133.19809
1	$\kappa_i(\gamma)$		7.78710e−1	5.43015e−1	8.71502e−1	2.20625
	$\widehat{\kappa}_{ni}(\gamma)$	10	7.77820e−1	5.42902e−1	8.71372e−1	2.20598
		15	7.78679e−1	5.43014e−1	8.71502e−1	2.20625
		20	7.78709e−1	5.43015e−1	8.71502e−1	2.20625
		25	7.78710e−1	5.43015e−1	8.71502e−1	2.20625
		30	7.78710e−1	5.43015e−1	8.71502e−1	2.20625
2	$\kappa_i(\gamma)$		2.02207e−1	4.57767e−2	2.39200e−2	1.97036e−2
	$\widehat{\kappa}_{ni}(\gamma)$	10	2.02200e−1	4.57766e−2	2.39199e−2	1.97036e−2
		15	2.02207e−1	4.57767e−2	2.39200e−2	1.97036e−2
		20	2.02207e−1	4.57767e−2	2.39200e−2	1.97036e−2
		25	2.02207e−1	4.57767e−2	2.39200e−2	1.97036e−2
		30	2.02207e−1	4.57767e−2	2.39200e−2	1.97036e−2

Table 9 Estimators of the two largest eigenvalues of kernel (19) obtained for $5 \leq n \leq 13$ and $\gamma = 1.5, 2, 3$ with the Rayleigh–Ritz method

	$\gamma = 1.5$		$\gamma = 2$		$\gamma = 3$	
n	$\widehat{\lambda}_1$	$\widehat{\lambda}_2$	$\widehat{\lambda}_1$	$\widehat{\lambda}_2$	$\widehat{\lambda}_1$	$\widehat{\lambda}_2$
5	3.367700e−1	1.860092e−1	8.767026e−2	5.795029e−2	1.388053e−2	1.271398e−2
6	3.378871e−1	1.860092e−1	8.769646e−2	5.795029e−2	1.388069e−2	1.271398e−2
7	3.378871e−1	1.861737e−1	8.769646e−2	5.795333e−2	1.388069e−2	1.271400e−2
8	3.380118e−1	1.861737e−1	8.769861e−2	5.795333e−2	1.388070e−2	1.271400e−2
9	3.380118e−1	1.861792e−1	8.769861e−2	5.795335e−2	1.388070e−2	1.271400e−2
10	3.380137e−1	1.861792e−1	8.769862e−2	5.795335e−2	1.388070e−2	1.271400e−2
11	3.380137e−1	1.861794e−1	8.769862e−2	5.795335e−2	1.388070e−2	1.271400e−2
12	3.380138e−1	1.861794e−1	8.769862e−2	5.795335e−2	1.388070e−2	1.271400e−2
13	3.380138e−1	1.861794e−1	8.769862e−2	5.795335e−2	1.388070e−2	1.271400e−2

$$\langle a, b\rangle = \sum_{t\geq 0} a_t \bar{b}_t w_\varrho(t),$$

for $a = (a_0, a_1, \ldots)$, $b = (b_0, b_1, \ldots) \in \ell^2(\varrho)$. The practical computation of this inner product involves truncation of the infinite sum, say from $t = 0$ up to $t = v$, for some positive integer v.

A suitable set of orthonormal polynomials can be expressed as

$$\phi_k(x;\varrho) = \left(\varrho^k\, k!\right)^{1/2} C_k(x;\varrho), \quad x \in \mathbb{N}_0,$$

where $C_k(x;\varrho) = \sum_{\nu=0}^{k} (-1)^{k-\nu} \binom{k}{\nu} \nu! \varrho^{-\nu} \binom{x}{\nu}$ represents the Charlier polynomial of degree k.

Example 4.4 (Testing for the von Mises distribution) The covariance kernel in Theorem 1 of Jammalamadaka et al. (2019) is

$$K(s,t;\theta) = \mathbb{E}\{\Upsilon(s,X;\theta)\overline{\Upsilon}(t,X;\theta)\}, \quad s,t \in \mathbb{N}_0, \tag{21}$$

where

$$\begin{aligned}\Upsilon(s,X;\theta) = &\cos(rX) - \Re\varphi(s;\theta) - \nabla\Re\varphi(s;\theta)^\top L(X;\theta) \\ &+ \mathrm{i}\left\{\sin(rX) - \Im\varphi(s;\theta) - \nabla\Im\varphi(s;\theta)^\top L(X;\theta)\right\},\end{aligned}$$

θ is the vector of parameters of the law in the null hypothesis, X is a circular random variable having the law in the null hypothesis with parameter vector θ, $\Re\varphi(s;\theta)$ ($\Im\varphi(s;\theta)$) is the real (imaginary) part of the characteristic function of X, $\nabla\Re\varphi(s;\theta)$ ($\nabla\Im\varphi(s;\theta)$) is the derivative of $\Re\varphi(s;\theta)$ ($\Im\varphi(s;\theta)$) with respect to θ, $L(X;\theta)$ is the linear term in the Bahadur expansion of the estimator of θ, and $\overline{\Upsilon}$ is the complex conjugate of Υ.

When applied to testing goodness-of-fit to the von Mises distribution, and the parameters μ (mean) and τ (concentration) are estimated with their maximum likelihood estimators, the above quantities become [see Mardia and Jupp (2000)]

$$\begin{aligned}\Re\varphi(s;\mu,\tau) &= \cos(\mu s)q(s;\tau), \\ \Im\varphi(s;\mu,\tau) &= \sin(\mu s)q(s,\tau), \\ L(X;\mu,\tau) &= \begin{pmatrix} \sin(x-\mu)/q(1;\tau) \\ \{\cos(x-\mu) - q(1;\tau)\}/\{1 - q(1;\tau)^2 - q(1;\tau)/\tau\} \end{pmatrix},\end{aligned}$$

where $q(s;\tau) = I_s(\tau)/I_0(\tau)$ and $I_s(\tau)$ denotes the modified Bessel of the first kind and order s [see, e.g. Chapter 9 of Olver et al. (2010)]. After some calculations, one gets

$$K(s,t) = \cos\{(s-t)\mu\}\, Q(s,t;\tau) + \mathrm{i}\sin\{(s-t)\mu\}\, Q(s,t;\tau), \quad s,t \in \mathbb{N}_0, \tag{22}$$

where

$$\begin{aligned}Q(s,t;\tau) = q(s-t;\tau) - q(s;\tau)q(t;\tau)\{1 + st/(\tau q(1;\tau)\} \\ - q'(s;\tau)q'(t,\tau)/\left\{1 - q(1;\tau)^2 - q(1;\tau)/\tau\right\}\end{aligned}$$

and $q'(s;\tau) = \frac{\partial}{\partial s} q(s;\tau)$.

Table 10 displays the estimators of the two largest eigenvalues obtained for $n = 10, 15, 20$, $v = 10$ (the same values were obtained for larger values of v) and $\mu = 0$. As in the previous examples, rapid convergence is observed.

We juxtapose the results in Table 10 with those obtained by implementing the Monte Carlo procedure outlined in Appendix A1. The Monte Carlo procedure was repeated 500 times, for various values of N (the sample sizes). Table 11 showcases the mean and the standard deviation of the two largest values across the 500 replications. The means in Table 11 approximate the values in Table 10, and they converge as N increases. The standard deviations in Table 11 diminish as N increases. From a computational perspective, the Rayleigh–Ritz method outperforms the Monte Carlo procedure, as the former computes the eigenvalues of a significantly smaller matrix.

4.5 Support $M = \mathbb{R}^d$

In this scenario, we concentrate on the weight function $w_\gamma(t) = \exp(-\gamma\|t\|^2)$, $t \in \mathbb{R}^d$, $\gamma > 0$, where $\|x\| = \sqrt{x^\top x}$ represents the Euclidean norm and $\top$ denotes the transpose of a vector. In the spirit of Baringhaus (1996), a suitable set of orthonormal polynomials can be expressed using (16) as

$$\widetilde{\phi}_{k_1,\dots,k_d}(x;\gamma) = \prod_{j=1}^{d} \phi_{k_j}(x_j) = \prod_{j=1}^{d} \left(2^{k_j}\, k_j! \sqrt{\pi/\gamma}\right)^{-d/2} H_{k_j}\left(\sqrt{\gamma} x_j\right), \qquad x = (x_1,\dots,x_d) \in \mathbb{R}^d, \quad k_1,\dots,k_d \in \mathbb{N}_0. \tag{23}$$

We focus our attention on the classic BHEP test as presented in Henze and Wagner (1997) with covariance kernel given in display (2.3)

$$K(s,t) = \exp\left(-\frac{\|s-t\|^2}{2}\right) - \left(1 + s^\top t + \frac{(s^\top t)^2}{2}\right)\exp\left(-\frac{\|s\|^2 + \|t\|^2}{2}\right), \quad s,t \in \mathbb{R}^d. \tag{24}$$

The weight function considered in Henze and Wagner (1997) is

$$w(t) = \varphi_\beta(t) = (2\pi\beta^2)^{-d/2} \exp\left(-\frac{\|t\|^2}{2\beta^2}\right), \quad t \in \mathbb{R}^d, \quad \beta > 0.$$

Clearly, for $\gamma = 1/(2\beta^2)$, we have that

Table 10 Estimators of the two largest eigenvalues of kernel (22) obtained for $n = 10, 15, 20$, $v = 10$ and $\mu = 0$, with the Rayleigh–Ritz method

ϱ	τ	n	$\widehat{\lambda}_1$	$\widehat{\lambda}_2$	ϱ	τ	n	$\widehat{\lambda}_1$	$\widehat{\lambda}_2$
0.5	1	10	6.288772e−2	9.849891e−3	0.5	5	10	8.438256e−3	5.328604e−4
		15	6.288772e−2	9.849896e−3			15	8.438256e−3	5.328604e−4
		20	6.288772e−2	9.849896e−3			20	8.438256e−3	5.328604e−4
1	1	10	1.632234e−1	4.686629e−2	1	5	10	3.826515e−2	3.445026e−3
		15	1.632234e−1	4.686841e−2			15	3.826515e−2	3.445029e−3
		20	1.632234e−1	4.686841e−2			20	3.826515e−2	3.445029e−3

Table 11 Mean and standard deviation (in parenthesis) in 500 replications of the Monte Carlo method for estimating the two largest eigenvalues of kernel (22), for several values of N

ϱ	τ	N	$\widehat{\lambda}_1$		$\widehat{\lambda}_2$		ϱ	τ	N	$\widehat{\lambda}_1$		$\widehat{\lambda}_2$	
0.5	1	50	6.349153e−2	(26.8e−3)	8.824334e−3	(10.2e−3)	0.5	5	50	9.168138e−3	(6.8e−3)	2.903251e−4	(5.3e−4)
		100	6.277916e−2	(19.0e−3)	9.430931e−3	(7.3e−3)			100	8.828349e−3	(4.9e−3)	3.828378e−4	(5.2e−4)
		250	6.313665e−2	(12.7e−3)	9.840565e−3	(4.6e−3)			250	8.721612e−3	(2.9e−3)	4.852314e−4	(4.3e−4)
		1000	6.286598e−2	(6.4e−3)	9.850643e−3	(2.5e−3)			1000	8.483449e−3	(1.4e−3)	5.301540e−4	(2.2e−4)
		2000	6.269159e−2	(4.5e−3)	9.840051e−3	(1.8e−3)			2000	8.538285e−3	(1.0e−3)	5.430429e−4	(1.6e−4)
		3000	6.270766e−2	(3.7e−3)	9.818919e−3	(1.5e−3)			3000	8.487909e−3	(1.0e−3)	5.405830e−4	(1.6e−4)
		4000	6.282082e−2	(3.6e−3)	9.825737e−3	(1.3e−3)			4000	8.471089e−3	(0.7e−3)	5.362364e−4	(1.1e−4)
		5000	6.288797e−2	(2.8e−3)	9.787226e−3	(1.1e−3)			5000	8.453944e−3	(0.6e−3)	5.351108e−4	(1-0e-4)
1	1	50	1.663794e−1	(43.0e−3)	4.383321e−2	(18.9e−3)	1	5	50	3.826183e−2	(13.9e−3)	2.835939e−3	(26.6e−4)
		100	1.648465e−1	(28.6e−3)	4.481114e−2	(13.6e−3)			100	3.836605e−2	(10.0e−3)	3.294036e−3	(21.8e−4)
		250	1.631288e−1	(17.6e−3)	4.576026e−2	(8.7e−3)			250	3.830483e−2	(6.7e−3)	3.339658e−3	(13.6e−4)
		1000	1.631152e−1	(9.3e−3)	4.652113e−2	(4.4e−3)			1000	3.813966e−2	(3.3e−3)	3.434499e−3	(6.4e−3)
		2000	1.629177e−1	(6.4e−3)	4.654794e−2	(3.2e−3)			2000	3.826935e−2	(2.4e−3)	3.452576e−3	(4.6e−3)
		3000	1.631266e−1	(5.5e−3)	4.661693e−2	(2.6e−3)			3000	3.832378e−2	(1.9e−3)	3.464659e−3	(3.9e−4)
		4000	1.631981e−1	(4.8e−3)	4.674310e−2	(2.1e−3)			4000	3.834800e−2	(1.7e−3)	3.462685e−3	(3.4e−4)
		5000	1.632090e−1	(4.3e−3)	4.678101e−2	(1.9e−3)			5000	3.832935e−2	(1.5e−3)	3.459746e−3	(3.1e−4)

$$w_\gamma(t) = (2\pi\beta^2)^{d/2}\varphi_\beta(t), \tag{25}$$

and in such a case, from (25),

$$\int K(s,t)f(s)\varphi_\beta(s)ds = \theta f(t)$$

if and only if

$$\int K(s,t)f(s)w_\gamma(s)ds = \lambda f(t), \quad \text{with} \quad \lambda = (2\pi\beta^2)^{d/2}\theta.$$

The exact expression of the first three cumulants of the law of the distribution of $T(\beta) = \|Z\|_{L^2}^2$ with $w = \varphi_\beta$ (denote them as $\kappa_1(\beta)$, $\kappa_2(\beta)$ and $\kappa_3(\beta)$) have been given in Henze and Wagner (1997). If such cumulants are denoted as $\kappa_1(\gamma)$, $\kappa_2(\gamma)$ and $\kappa_3(\gamma)$, respectively, when $w = w_\gamma$, then from (25), for $\gamma = 1/(2\beta^2)$, we have that

$$\kappa_1(\gamma) = (\pi/\gamma)^{d/2}\kappa_1(\beta), \quad \kappa_2(\gamma) = (\pi/\gamma)^d\kappa_2(\beta), \quad \kappa_3(\gamma) = (\pi/\gamma)^{3d/2}\kappa_3(\beta).$$

In order to check the accuracy of the solutions of Eq. (2), we calculated the exact values of $\kappa_1(\gamma)$, $\kappa_2(\gamma)$ and $\kappa_3(\gamma)$ and the approximations in (18), where now *m* is the number of non-null solutions of Eq. (2) obtained by considering all elements in the basis (23) with $k_1 + \ldots + k_d \le n$. Tables 12, 13 and 14 display $\kappa_i(\gamma)$ and $\widehat{\kappa}_{ni}(\gamma)$ for $\gamma = 0.5, 1, 2$, $n = 10(5)30$ and $d = 1, 2, 3$. Looking at these tables it can be concluded that $\kappa_i(\gamma)$ and $\widehat{\kappa}_{ni}(\gamma)$ are very close for $n \ge 15$, specially when $\gamma = 1, 2$.

Note that using the formulas from the first point in Example 4.3 we can easily approximate the hitherto unknown fourth cumulant of the limit distribution of the BHEP test. This would allow to fit a Pearson system of distributions as described in Sect. 6.1 to efficiently approximate the critical values of the test statistic.

5 Covariance kernels depending on unknown parameters

Until now, we have assumed that the covariance kernel is explicitly known. This assumption holds for all the examples presented in Sect. 4, with the exception of Example 4.4. In that particular case, the kernel depends on two parameters. To obtain the approximations shown in Tables 10 and 11, we fixed the values $\mu = 0$ and $\tau = 1,\ 5$. As outlined in the Introduction, a central motivation for approximating the eigenvalues of a kernel lies in its application to goodness-of-fit testing. For location-scale families, the kernel associated with the test statistic typically does not depend on unknown parameters. In contrast, for families involving a shape parameter, the kernel generally depends on this parameter, which is often unknown in practice. If the kernel varies continuously with respect to the parameter, the result stated in Theorem 3.1 remains valid when the parameter is substituted by a consistent estimator. This section establishes this result and presents two illustrative examples.

Table 12 First three cumulants and their Rayleigh–Ritz estimators of the distribution of $T(\gamma) = \|Z\|^2_{L^2}$, Z a centered Gaussian process with covariance kernel K in (24) and $d = 1$

γ		n	$i = 1$	$i = 2$	$i = 3$
0.5	$\kappa_i(\gamma)$	True	3.35825e−1	9.57324e−2	6.30525e–2
	$\hat{\kappa}_{ni}(\gamma)$	10	3.30722e−1	9.41680e−2	6.16292e–2
		15	3.35124e−1	9.56358e−2	6.29862e–2
		20	3.35728e−1	9.57275e−2	6.30492e–2
		25	3.35813e−1	9.57322e−2	6.30524e–2
		30	3.35823e−1	9.57324e−2	6.30525e–2
1	$\kappa_i(\gamma)$	True	8.83130e−2	8.07001e−3	1.69471e–3
	$\hat{\kappa}_{ni}(\gamma)$	10	8.81463e−2	8.05925e−3	1.69165e–3
		15	8.83081e−2	8.06992e−3	1.69469e–3
		20	8.83128e−2	8.07001e−3	1.69471e–3
		25	8.83130e−2	8.07001e−3	1.69471e–3
		30	8.83130e−2	8.07001e−3	1.69471e–3
2	$\kappa_i(\gamma)$	True	1.67944	3.51349e−4	1.68584e–5
	$\hat{\kappa}_{ni}(\gamma)$	10	1.67930	3.51336e−4	1.68575e–5
		15	1.67944	3.51349e−4	1.68584e–5
		20	1.67944	3.51349e−4	1.68584e–5
		25	1.67944	3.51349e−4	1.68584e–5
		30	1.67944	3.51349e−4	1.68584e–5

Table 13 First three cumulants and their Rayleigh–Ritz estimators of the distribution of $T(\gamma) = \|Z\|^2_{L^2}$, Z a centered Gaussian process with covariance kernel K in (24) and $d = 2$

γ		n	$i = 1$	$i = 2$	$i = 3$
0.5	$\kappa_i(\gamma)$	True	1.86169	7.38190e−1	7.87273e−1
	$\widehat{\kappa}_{ni}(\gamma)$	10	1.80572	7.24167e−1	7.67663e−1
		15	1.85336	7.37210e−1	7.86174e−1
		20	1.86047	7.38133e−1	7.87209e−1
		25	1.86152	7.38187e−1	7.87270e−1
		30	1.86166	7.38190e−1	7.87272e−1
1	$\kappa_i(\gamma)$	True	3.92699e−1	4.36479e−2	1.28019e−2
	$\widehat{\kappa}_{ni}(\gamma)$	10	3.91340e−1	4.35734e−2	1.27726e−2
		15	3.92654e−1	4.36472e−2	1.28018e−2
		20	3.92698e−1	4.36479e−2	1.28019e−2
		25	3.92699e−1	4.36479e−2	1.28019e−2
		30	3.92699e−1	4.36479e−2	1.28019e−2
2	$\kappa_i(\gamma)$	True	5.81776e−2	1.21539e−3	6.65175e−5
	$\widehat{\kappa}_{ni}(\gamma)$	10	5.81692e−2	1.21533e−3	6.65130e−5
		15	5.81776e−2	1.21539e−3	6.65175e−5
		20	5.81776e−2	1.21539e−3	6.65175e−5
		25	5.81776e−2	1.21539e−3	6.65175e−5
		30	5.81776e−2	1.21539e−3	6.65175e−5

Table 14 First three cumulants and their Rayleigh–Ritz estimators of the distribution of $T(\gamma) = \|Z\|^2_{L^2}$, Z a centered Gaussian process with covariance kernel K in (24) and $d = 3$

γ		n	$i = 1$	$i = 2$	$i = 3$
0.5	$\kappa_i(\gamma)$	True	7.16174	3.89456	6.64335
	$\widehat{\kappa}_{ni}(\gamma)$	10	6.78990	3.80774	6.46828
		15	7.09987	3.88788	6.63171
		20	7.15194	3.89413	6.64257
		25	7.16030	3.89453	6.64331
		30	7.16153	3.89456	6.64335
1	$\kappa_i(\gamma)$	True	1.20027	1.61360e−1	6.53432e−2
	$\widehat{\kappa}_{ni}(\gamma)$	10	1.19318	1.61016e−1	6.51617e−2
		15	1.20027	1.61356e−1	6.53418e−2
		20	1.20026	1.61360e−1	6.53432e−2
		25	1.20027	1.61360e−1	6.53432e−2
		30	1.20027	1.61360e−1	6.53432e−2
2	$\kappa_i(\gamma)$	true	1.38008e−1	2.87382e−3	1.77339e−4
	$\widehat{\kappa}_{ni}(\gamma)$	10	1.37974e−1	2.87363e−3	1.77323e−4
		15	1.38008e−1	2.87382e−3	1.77339e−4
		20	1.38008e−1	2.87382e−3	1.77339e−4
		25	1.38008e−1	2.87382e−3	1.77339e−4
		30	1.38008e−1	2.87382e−3	1.77339e−4

Throughout this section, we assume that the covariance kernel is known up to a finite-dimensional vector of unknown parameters θ, i.e., $K(s, r) = K(s, r; \theta)$ for $s, r \in M$, where $\theta \in \Theta \subseteq \mathbb{R}^d$ for some fixed $d \in \mathbb{N}$. Note that the eigenvalues of $K(s, r; \theta)$ also depend on θ, and we denote them by $\lambda_i(\theta)$. In most practical scenarios, the parameter θ is unknown. Therefore, to apply the Rayleigh–Ritz method in Sect. 3, it is necessary to substitute θ in the expression of K with a consistent estimator.

Let $\{\psi_j\}_{j\geq 1}$ be as defined in Sect. 3. Next, let $X_1, \ldots, X_N$ denote a random sample, and let $\hat{\theta} = \hat{\theta}(X_1, \ldots, X_N)$ be an estimator of the parameter θ. Define $K_n(s, t; \theta) = \sum_{j,k=1}^n K_{jk}(\theta)\psi_j(s)\psi_k(t)$ as the analogue of (3), where $K_{jk}(\theta) = \langle \mathcal{K}_\theta \psi_j, \psi_k \rangle = \langle \psi_j, \mathcal{K}_\theta \psi_k \rangle$, and $\mathcal{K}_\theta$ denotes the operator associated with the covariance kernel $K(\cdot, \cdot; \theta)$. Let $M_n(\theta) = (K_{jk}(\theta))_{1\leq j,k\leq n}$ be the $n \times n$ matrix of coefficients. The following result shows that, under mild regularity conditions, the eigenvalues of $M_n(\hat{\theta})$ provide a good approximation to the eigenvalues of the kernel $K(\cdot, \cdot; \theta)$.

Theorem 5.1 *Let $\{\lambda_i(\theta)\}_{i\geq 1}$ denote the eigenvalues of $K(\cdot, \cdot; \theta)$, arranged in decreasing order and repeated according to their multiplicity. Let $\widehat{\lambda}_1(\hat{\theta}) \geq \widehat{\lambda}_2(\hat{\theta}) \geq \ldots \geq \widehat{\lambda}_n(\hat{\theta})$ be the eigenvalues of the matrix $M_n(\hat{\theta})$. Suppose that $\hat{\theta} \xrightarrow{a.s.(P)} \theta$ and that $\|K(\cdot, \cdot; \theta) - K(\cdot, \cdot; \gamma)\|_{\mathcal{H}} \to 0$ as $\|\theta - \gamma\| \to 0$. Then, $|\widehat{\lambda}_i - \lambda_i| \xrightarrow{a.s.(P)} 0$, as $N, n \to \infty$, for each i.*

Proof We have that

$$\|K_n(\cdot,\cdot;\theta)-K_n(\cdot,\cdot;\hat{\theta})\|_{\mathcal{H}}^2=\int_M\int_M\left\{K_n(s,t;\theta)-K_n(s,t;\hat{\theta})\right\}^2\mu(\mathrm{d}s)\mu(\mathrm{d}t)$$
$$=\sum_{j,k,u,v=1}^{n}\left\{K_{jk}(\theta)-K_{jk}(\hat{\theta})\right\}\left\{K_{uv}(\theta)-K_{uv}(\hat{\theta})\right\}\int_M\psi_j(s)\psi_u(s)\mu(\mathrm{d}s)\int_M\psi_k(t)\psi_v(t)\mu(\mathrm{d}t)$$
$$=\sum_{j,k=1}^{n}\left\{K_{jk}(\theta)-K_{jk}(\hat{\theta})\right\}^2\le\sum_{j,k\ge1}\left\{K_{jk}(\theta)-K_{jk}(\hat{\theta})\right\}^2=\|K(\cdot,\cdot;\theta)-K(\cdot,\cdot;\hat{\theta})\|_{\mathcal{H}}^2,$$

where the last equality comes from Parseval identity. The assumptions made imply that $\|K(\cdot,\cdot;\theta)-K(\cdot,\cdot;\hat{\theta})\|_{\mathcal{H}}\xrightarrow{a.s.(P)}0$, as $N\to\infty$, and hence,

$$\|K_n(\cdot,\cdot;\theta)-K_n(\cdot,\cdot;\hat{\theta})\|_{\mathcal{H}}\xrightarrow{a.s.(P)}0,\quad N\to\infty. \tag{26}$$

Finally, taking into account that [recall (6)]

$$\|K(\cdot,\cdot;\theta)-K_n(\cdot,\cdot;\hat{\theta})\|_{\mathcal{L}}\le\|K(\cdot,\cdot;\theta)-K_n(\cdot,\cdot;\theta)\|_{\mathcal{L}}+\|K_n(\cdot,\cdot;\theta)-K_n(\cdot,\cdot;\hat{\theta})\|_{\mathcal{L}}$$
$$\le\|K(\cdot,\cdot;\theta)-K_n(\cdot,\cdot;\theta)\|_{\mathcal{H}}+\|K_n(\cdot,\cdot;\theta)-K_n(\cdot,\cdot;\hat{\theta})\|_{\mathcal{H}},$$

the desired result follows from the above inequality, (4), (5) and (26). □

As in the case where K is known, one can similarly obtain estimators (up to sign) of the eigenfunctions of $K(\cdot,\cdot;\theta)$. Furthermore, the proof of Theorem 3.3 can be adapted to demonstrate that these estimators converge (up to sign) to the true eigenfunctions as $N,n\to\infty$.

Example 5.2 (Poissonity Test) The covariance kernel associated with the statistic used for testing the goodness-of-fit to the Poisson distribution, as introduced in Rueda and O'Reilly (1999); Rueda et al. (1991), is given by

$$K(s,t;\theta)=e^{\theta(st-1)}-e^{\theta(s+t-2)}-\theta(s-1)(t-1)e^{\theta(s+t-2)},\quad 0\le s,t\le 1, \tag{27}$$

where θ denotes the (unknown) mean of the Poisson distribution. Since $M=[0,1]$ and $w(t)=1$ we employ the basis introduced in Sect. 4.1. Table 15 reports the means and standard deviations (in parentheses) over 1000 replications of the three largest eigenvalues of the kernel (27), where θ is replaced by its sample estimate $\hat{\theta}$ (the sample mean of N observations drawn from a Poisson distribution with mean $\theta=1$). Results are shown for finite values of N as well as for the limiting case $\theta=1$ when $N=\infty$, across various values of n and N, using the Rayleigh–Ritz method. Figure 3 displays these eigenvalues. As in the case when the kernel is fully known, small values of n yield good approximations to the largest eigenvalues. Consistent with Theorem 5.1, the eigenvalues of $K(s,t;\hat{\theta})$ approach those of $K(s,t;\theta)$ as N increases.

Example 5.3 We revisit the kernel introduced in Example 4.4, this time addressing the scenario in which the parameters are unknown and must be estimated from the

Table 15 Mean and standard deviation (in parenthesis) in 1000 replications of the three largest eigenvalues of kernel (27) with θ replaced with $\hat{\theta}$ (the sample mean of N data generated from a Poisson population with $\theta = 1$) for finite N and $\theta = 1$ for $N = \infty$, for several values of n and N, with the Rayleigh–Ritz method

n	N	$\widehat{\lambda}_1(\hat{\theta})$		$\widehat{\lambda}_2(\hat{\theta})$		$\widehat{\lambda}_3(\hat{\theta})$		n	N	$\widehat{\lambda}_1(\hat{\theta})$		$\widehat{\lambda}_2(\hat{\theta})$		$\widehat{\lambda}_3(\hat{\theta})$	
3	50	2.432e−2	(23.4e−4)	2.482e−4	(71.1e−6)	2.395e−6	(10.9e−7)	5	50	2.432e−2	(23.4e−4)	2.486e−4	(71.4e−6)	2.624e−6	(11.6e−7)
	100	2.444e−2	(16.0e−4)	2.439e−4	(48.5e−6)	2.250e−6	(4.6e−7)		100	2.444e−2	(16.0e−4)	2.443e−4	(48.7e−6)	2.474e−6	(4.9e−7)
	250	2.458e−2	(10.1e−4)	2.441e−4	(31.4e−6)	2.250e−6	(4.6e−7)		250	2.458e−2	(10.1e−4)	2.446e−4	(31.5e−6)	2.474e−6	(4.9e−7)
	500	2.465e−2	(7.1e−4)	2.448e−4	(22.2e−6)	2.249e−6	(3.2e−7)		500	2.465e−2	(7.1e−4)	2.452e−4	(22.2e−6)	2.474e−6	(3.5e−7)
	1000	2.465e−2	(5.0e−4)	2.441e−4	(15.7e−6)	2.234e−6	(2.3e−7)		1000	2.465e−2	(5.0e−4)	2.445e−4	(15.7e−6)	2.457e−6	(2.5e−7)
	2000	2.467e−2	(3.4e−4)	2.442e−4	(10.6e−6)	2.232e−6	(1.6e−7)		2000	2.467e−2	(3.4e−4)	2.446e−4	(10.8e−6)	2.456e−6	(1.7e−7)
	3000	2.466e−2	(2.8e−4)	2.440e−4	(8.7e−6)	2.229e−6	(1.3e−7)		3000	2.466e−2	(2.8e−4)	2.444e−4	(8.8e−6)	2.452e−6	(1.4e−7)
	4000	2.467e−2	(2.5e−4)	2.442e−4	(7.7e−6)	2.232e−6	(1.1e−7)		4000	2.467e−2	(2.5e−4)	2.446e−4	(7.7e−6)	2.455e−6	(1.2e−7)
	5000	2.467e−2	(2.2e−4)	2.441e−4	(6.7e−6)	2.230e−6	(1.0e−7)		5000	2.467e−2	(2.2e−4)	2.442e−4	(6.7e−6)	2.453e−6	(1.1e−7)
	∞	2.467e−2		2.439e−4		2.226e−6			∞	2.467e−2		2.443e−4		2.449e−6	
4	50	2.432e−2	(23.4e−4)	2.486e−4	(71.4e−6)	2.616e−6	(11.6e−7)	6	50	2.432e−2	(23.4e−4)	2.486e−4	(71.4e−6)	2.625e−6	(11.6e−7)
	100	2.444e−2	(16.0e−4)	2.443e−4	(48.7e−6)	2.467e−6	(4.9e−7)		100	2.444e−2	(16.0e−4)	2.443e−4	(48.7e−6)	2.474e−6	(4.9e−7)
	250	2.458e−2	(10.1e−4)	2.446e−4	(31.5e−6)	2.467e−6	(4.9e−7)		250	2.458e−2	(10.1e−4)	2.446e−4	(31.5e−6)	2.474e−6	(4.9e−7)
	500	2.465e−2	(7.1e−4)	2.452e−4	(22.3e−6)	2.467e−6	(3.5e−7)		500	2.465e−2	(7.1e−4)	2.452e−4	(22.3e−6)	2.474e−6	(3.5e−7)
	1000	2.465e−2	(5.0e−4)	2.445e−4	(15.7e−6)	2.451e−6	(2.5e−7)		1000	2.465e−2	(5.0e−4)	2.445e−4	(15.7e−6)	2.457e−6	(2.5e−7)
	2000	2.466e−2	(3.4e−4)	2.446e−4	(10.8e−5)	2.449e−6	(1.7e−7)		2000	2.466e−2	(3.4e−4)	2.446e−4	(10.8e−6)	2.456e−6	(1.7e−7)
	3000	2.466e−2	(2.8e−4)	2.444e−4	(8.8e−6)	2.445e−6	(1.4e−7)		3000	2.466e−2	(2.8e−4)	2.444e−4	(8.8e−6)	2.452e−6	(1.4e−7)
	4000	2.467e−2	(2.5e−4)	2.446e−4	(7.7e−6)	2.449e−6	(1.2e−7)		4000	2.467e−2	(2.5e−4)	2.446e−4	(7.7e−6)	2.456e−6	(1.2e−7)
	5000	2.467e−2	(2.2e−4)	2.442e−4	(6.7e−6)	2.446e−6	(1.1e−7)		5000	2.467e−2	(2.2e−4)	2.442e−4	(6.7e−6)	2.453e−6	(1.1e−7)
	∞	2.467e−2		2.443e−4		2.443e−6			∞	2.467e−2		2.443e−4		2.449e−6	

observed data. Table 16 parallels Table 10, but with θ replaced by its maximum likelihood estimator, $\hat{\theta}$. Specifically, Table 16 reports the mean and standard deviation (in parentheses) of the two largest eigenvalues of the kernel in Eq. (27), computed using the Rayleigh–Ritz method, across 1000 replications. Results are presented for several combinations of n and N, with θ substituted by $\hat{\theta}$. As can be seen from a one-to-one comparison of the two tables, the eigenvalues differ slightly. This discrepancy may be attributed to the additional layer of randomness introduced by using the MLE estimator $\hat{\theta}$.

6 Usefulness of eigenvalues estimation

In this section, we introduce two statistical use cases where an accurate approximation of the eigenvalues of the covariance operator is of paramount importance.

6.1 Limit distribution approximation

As noted in the Introduction, if Z is a Gaussian process that takes values in an appropriate function space and $\|\cdot\|$ is a corresponding norm, then $\|Z\|^2$ follows the distribution $W = \sum_{j=1}^{\infty} \lambda_j N_j^2$, where $N_1, N_2, \ldots$ are iid standard normal random variables, and $\{\lambda_j\}_{j \in \mathbb{N}}$ is the sequence of positive eigenvalues of the covariance operator, $\mathcal{K}$, of Z. Therefore, if $\widehat{\lambda}_1, \ldots, \widehat{\lambda}_m$ are estimators of the m largest eigenvalues of $\mathcal{K}$, the distribution of $\|Z\|^2$ can be approximated by that of $\widehat{W}_m = \sum_{j=1}^{m} \widehat{\lambda}_j N_j^2$. The distribution of a linear combination of chi-squared variates can subsequently be approximated by simulation or by employing the method of Imhof (1961), which is implemented in the `R` package `CompQuadForm` Duchesne and de Micheaux (2010). This procedure has been utilized in Novoa-Muñoz and Jiménez-Gamero (2016); Rivas Martínez and Jiménez-Gamero (2021); Rivas-Martínez et al. (2019) with eigenvalue estimators obtained as described in Appendix A1.

Next, we observe that by approximating the eigenvalues of the covariance operator, we can approximate the limiting distribution by a Pearson system [with the help of the `R` package `PearsonDS` Becker and Klüßner (2024)]. Initially, by directly evaluating integrals, the first two cumulants of the distribution of $\|Z\|_{L^2}^2$ are

$$\kappa_1 = \mathbb{E}\|Z\|_{L^2}^2 = \int_M K(t,t)\, w(t)\ \mathrm{d}\,t$$

and

$$\kappa_2 = \mathbb{V}(\|Z\|_{L^2}^2) = 2\int_M \int_M K^2(s,t)\, w(s)w(t)\ \mathrm{d}\,s\ \mathrm{d}\,t,$$

where $\mathbb{V}(\cdot)$ represents the variance. Following the methodology in Henze (1990); Stephens (1976), the third and fourth cumulants can be computed as

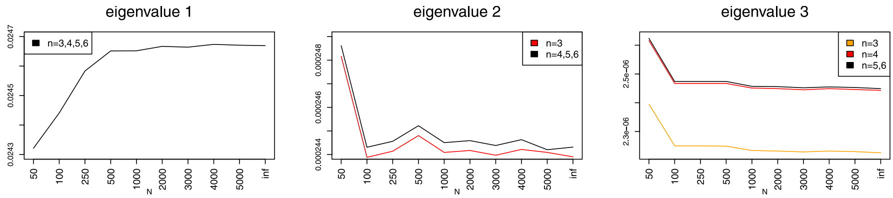

Fig. 3 Mean in 1000 replications of the three largest eigenvalues of kernel (27) with θ replaced with $\hat{\theta}$ (the sample mean of N data generated from a Poisson population with $\theta = 1$) for finite N and $\theta = 1$ for $N = \infty$, for several values of n and N, with the Rayleigh–Ritz method

$$\kappa_j = 2^{j-1}(j-1)! \int_M K_j(t,t)w(t) \ \mathrm{d}\ t,$$

where $K_j(s,t)$, the j^{th} iterate of $K(s, t)$, is defined as

$$\begin{aligned} K_j(s,t) =& \int_M K_{j-1}(s,u)K_0(u,t)w(u) \ \mathrm{d}\ u, \quad j \geq 2, \\ K_1(s,t) =& K(s,t). \end{aligned}$$

Utilizing the identities in Remark 2.2 of Ebner (2023) [also compare to Corollary 1.3 in Deheuvels and Martynov (2003)], we observe that

$$\sum_{k=1}^{\infty} \lambda_k^j = \kappa_j/(2^{j-1}(j-1)!), \quad j = 1,2,3,4, \tag{28}$$

thus sums of powers of the eigenvalues approximate the cumulants of the limiting distribution. This approach, applied by direct calculation of $\kappa_1, \kappa_2, \kappa_3, \kappa_4$ as described above, has been utilized in Ebner (2020); Ebner and Henze (2023); Henze (1990).

Example 6.1 We address the problem of testing the goodness-of-fit to a fully specified continuous univariate distribution using the Cramér–von Mises test statistic. In this context, the null distribution of the test statistic is free from unknown parameters and can therefore be computed (at least via simulation) for each sample size N. To evaluate the applicability of the aforementioned approximations for estimating the null distribution, we conducted the following experiment: a sample of size N was drawn from the uniform distribution on the unit interval, the Cramér–von Mises test statistic was computed, and the corresponding p-value was approximated using the following methods: the distribution of $\widehat{W}_m$ with $(n = 10,\ m = 5)$, $(n = m = 10)$, $(n = 20,\ m = 10)$, $(n = m = 20)$; the distribution of $W_\ell = \sum_{j=1}^{\ell} \lambda_j N_j^2$ with $\ell = 50,\ 100$; and the Pearson system with cumulants estimated under the same four combinations of n and m, and computed according to Eq. (28), with the infinite sum truncated at $\ell = 50,\ 100$. This procedure was repeated 10,000 times. Table 17 reports the proportion of p-values less than or equal to 0.05. From this table, we observe the following: the p-values obtained via the Pearson sys-

Table 16 Mean and standard deviation (in parenthesis) in 1000 replications of the two largest eigenvalues of kernel (21) with $\theta = (\mu, \tau)$ replaced with its maximum likelihood estimator, calculated from a sample of size N from the von Mises distribution with parameters $\mu = 0$ and τ as indicated in the table, using the Rayleigh–Ritz method, for several values of N and n

n	ϱ	τ	N	$\widehat{\lambda}_1$		$\widehat{\lambda}_2$		ϱ	τ	N	$\widehat{\lambda}_1$		$\widehat{\lambda}_2$	
10	0.5	1	50	6.168326e−2	(5.5e−3)	9.611427e−3	(11.5e−4)	0.5	5	50	8.522916e−3	(35.0e−4)	5.774242e−4	(3.6e−4)
			100	6.228327e−2	(3.8e−3)	9.729341e−3	(7.9e−4)			100	8.463742e−3	(24.1e−4)	5.530905e−4	(2.4e−4)
			250	6.281133e−2	(2.3e−3)	9.836178e−3	(4.9e−4)			250	8.528868e−3	(15.4e−4)	5.490065e−4	(1.5e−4)
			500	6.280986e−2	(1.6e−3)	9.834652e−3	(3.4e−4)			500	8.482871e−3	(10.7e−4)	5.407701e−4	(1.0e−4)
			1000	6.288746e−2	(1.2e−3)	9.850446e−3	(2.5e−4)			1000	8.506042e−3	(8.0e−4)	5.414898e−4	(0.8e−4)
			2000	6.297946e−2	(0.8e−3)	9.869482e−3	(1.7e−4)			2000	8.521136e−3	(5.4e−4)	5.418470e−4	(0.5e−4)
			3000	6.296019e−2	(0.7e−3)	9.865336e−3	(1.4e−4)			3000	8.487817e−3	(4.5e−4)	5.383112e−4	(0.4e−4)
			4000	6.297657e−2	(0.6e−3)	9.868727e−3	(1.2e−4)			4000	8.494489e−3	(3.9e−4)	5.388293e−4	(0.4e−4)
			5000	6.298205e−2	(0.5e−3)	9.869849e−3	(1.1e−4)			5000	8.491891e−3	(3.5e−4)	5.384802e−4	(0.3e−4)
	1	1	50	1.609481e−1	(9.6e−3)	4.574604e−2	(5.7e−3)	1	5	50	3.782999e−2	(12.2e−3)	3.583119e−3	(18.7e−4)
			100	1.620948e−1	(6.5e−3)	4.629663e−2	(3.9e−3)			100	3.799902e−2	(8.5e−3)	3.503421e−3	(12.7e−4)
			250	1.630558e−1	(3.9e−3)	4.681063e−2	(2.4e−3)			250	3.844019e−2	(5.4e−3)	3.510405e−3	(8.0e−4)
			500	1.630732e−1	(2.7e−3)	4.679623e−2	(1.7e−3)			500	3.835204e−2	(3.8e−3)	3.477091e−3	(5.6e−4)
			1000	1.632130e−1	(2.0e−3)	4.687252e−2	(1.2e−3)			1000	3.846471e−2	(2.8e−3)	3.485236e−3	(4.2e−4)
			2000	1.633737e−1	(1.4e−3)	4.696627e−2	(0.9e−3)			2000	3.854023e−2	(1.9e−3)	3.490273e−3	(2.8e−4)
			3000	1.633427e−1	(1.1e−3)	4.694494e−2	(0.7e−3)			3000	3.842811e−2	(1.6e−3)	3.472275e−3	(2.3e−4)
			4000	1.633712e−1	(1.0e−3)	4.696164e−2	(0.6e−3)			4000	3.845432e−2	(1.4e−3)	3.475404e−3	(2.0e−4)
			5000	1.633810e−1	(0.9e−3)	4.696710e−2	(0.5e−3)			5000	3.844704e−2	(1.2e−3)	3.473818e−3	(1.8e−4)
15	0.5	1	50	6.168326e−2	(5.5e−3)	9.611434e−3	(11.5e−4)	0.5	5	50	8.522916e−3	(35.0e−4)	5.774242e−4	(3.6e−4)
			100	6.228327e−2	(3.8e−3)	9.729347e−3	(7.9e−4)			100	8.463742e−3	(24.1e−4)	5.530905e−4	(2.4e−4)
			250	6.281133e−2	(2.3e−3)	9.836184e−3	(4.9e−4)			250	8.528868e−3	(15.4e−4)	5.490065e−4	(1.5e−4)
			500	6.280986e−2	(1.6e−3)	9.834657e−3	(3.4e−4)			500	8.482871e−3	(10.7e−4)	5.407701e−4	(1.0e−4)
			1000	6.288746e−2	(1.2e−3)	9.850452e−3	(2.5e−4)			1000	8.506042e−3	(8.0e−4)	5.414898e−4	(0.8e−4)
			2000	6.297946e−2	(0.8e−3)	9.869487e−3	(1.7e−4)			2000	8.521136e−3	(5.4e−4)	5.418470e−4	(0.5e−4)
			3000	6.296019e−2	(0.7e−3)	9.865341e−3	(1.4e−4)			3000	8.487817e−3	(4.5e−4)	5.383112e−4	(0.4e−4)
			4000	6.297657e−2	(0.6e−3)	9.868732e−3	(1.2e−4)			4000	8.494489e−3	(3.9e−4)	5.388293e−4	(0.4e−4)

Table 16 (continued)

n	ϱ	τ	N	$\widehat{\lambda}_1$		$\widehat{\lambda}_2$		ϱ	τ	N	$\widehat{\lambda}_1$		$\widehat{\lambda}_2$	
			5000	6.298205e−2	(0.5e−3)	9.869854e−3	(1.1e−4)			5000	8.491891e−3	(3.5e−4)	5.384802e−4	(0.3e−4)
	1	1	50	1.609482e−1	(9.6e−3)	4.574931e−2	(5.7e−3)	1	5	50	3.782999e−2	(12.2e−3)	3.583122e−3	(18.7e−4)
			100	1.620948e−1	(6.5e−3)	4.629931e−2	(3.9e−3)			100	3.799902e−2	(8.5e−3)	3.503424e−3	(12.7e−4)
			250	1.630558e−1	(3.9e−3)	4.681306e−2	(2.4e−3)			250	3.844019e−2	(5.4e−3)	3.510408e−3	(8.0e−4)
			500	1.630732e−1	(2.7e−3)	4.679848e−2	(1.7e−3)			500	3.835204e−2	(3.8e−3)	3.477094e−3	(5.6e−4)
			1000	1.632130e−1	(2.0e−3)	4.687474e−2	(1.2e−3)			1000	3.846471e−2	(2.8e−3)	3.485239e−3	(4.2e−4)
			2000	1.633737e−1	(1.4e−3)	4.696848e−2	(0.9e−3)			2000	3.854023e−2	(1.9e−3)	3.490276e−3	(2.8e−4)
			3000	1.633427e−1	(1.1e−3)	4.694713e−2	(0.7e−3)			3000	3.842811e−2	(1.6e−3)	3.472278e−3	(2.3e−4)
			4000	1.633713e−1	(1.0e−3)	4.696383e−2	(0.6e−3)			4000	3.845432e−2	(1.4e−3)	3.475408e−3	(2.0e−4)
			5000	1.633810e−1	(0.9e−3)	4.696929e−2	(0.5e−3)			5000	3.844704e−2	(1.2e−3)	3.473821e−3	(1.8e−4)

tem tend to be slightly smaller than those computed using Imhof's method; setting $n = m$ yields results closer to the nominal level of 0.05; as either n or ℓ increases, the p-values converge toward the target level. Overall, the results in the table are quite close to the nominal value of 0.05, particularly for larger values of N. This behavior is expected, as all the considered approximations aim to estimate the asymptotic null distribution. Moreover, according to Stephens (1970), the critical values for finite N differ from those of the asymptotic distribution by an amount of order $1/N$.

Example 6.2 We continue the analysis of the goodness-of-fit test for the Poisson distribution presented in Example 5.2. Since the null distribution of the test statistic is unknown, it is typically approximated using a parametric bootstrap (see Gürtler and Henze (2000) for results on the consistency of the bootstrap estimator of the null distribution). To evaluate the accuracy of the previous approximations to the null distribution and to compare them with the bootstrap estimator, we conducted the following experiment: A sample of size N was generated from a Poisson distribution with mean 1. The test statistic proposed in Rueda et al. (1991); Rueda and O'Reilly (1999) was computed, and the corresponding p-value was approximated using the following methods: (i) the parametric bootstrap described in Gürtler and Henze (2000) with 1000 bootstrap replications, (ii) the distribution of $\widehat{W}_m$ for $m = n = 9$ and $n = 9$, $m = 5$, and (iii) the Pearson system with cumulants estimated for $m = n = 9$ and $n = 9$, $m = 5$. This procedure was repeated 10,000 times. Table 18 reports the proportion of p-values less than or equal to 0.05. From this table, we observe the following: the results for $m = n = 5$ and $n = 9$, $m = 5$ are identical, indicating that omitting the smallest eigenvalues does not affect the p-value when using either $\widehat{W}_m$ or the Pearson system. Moreover, there is virtually no difference between these two approximations. All values in the table are quite close to the nominal level of 0.05, particularly as the sample size N increases. This outcome is expected for the $\widehat{W}_m$ and Pearson-based approximations, since both aim to estimate the asymptotic null distribution. For the bootstrap method, this behavior is also reasonable, as it yields a consistent estimator, providing results that converge to the nominal level with increasing sample size.

6.2 Bahadur efficiency

Here we demonstrate the importance of the approximation of eigenvalues for the test's quality assessment, by calculating local Bahadur efficiency - one of often-used asymptotic criteria. Its widespread adoption as an asymptotic quality measure is primarily due to its applicability to test statistics with non-normal limiting distributions. For details, we refer to Nikitin (1995), while a brief review is given in Appendix B.

Let $\mathcal{G} = \{G_\theta(x),\ \theta > 0\}$ be a family of alternative distribution functions, with $G_0(x)$ being the null family of distributions. The relative Bahadur efficiency of two sequences of test statistics $\{T_n\}$ and $\{V_n\}$ can be represented as the ratio of their approximate Bahadur slope functions $c_T^*(\theta)$ and $c_V^*(\theta)$, which are associated with the rate of exponential decrease for the level of significance achieved under the alternative. Typically, we are interested in the case when $\theta \to 0$, i.e., when we are comparing the behavior of tests against nearby alternatives.

Table 17 Empirical levels for testing goodness-of-fit to a continuous univariate distribution with the Cramér-von Mises test statistic, calculated with 10,000 replications, using $\widehat{W}_m$, W_ℓ (Imhof) and the Pearson System (Pearson), for several sample sizes (N)

	Imhof						Pearson					
	$n=10$	$n=10$	$n=20$	$n=20$			$n=10$	$n=10$	$n=20$	$n=20$		
N	$m=5$	$m=10$	$m=10$	$m=20$	$\ell=50$	$\ell=100$	$m=5$	$m=10$	$m=10$	$m=20$	$\ell=50$	$\ell=100$
10	0.0538	0.0523	0.0519	0.0507	0.0495	0.0490	0.0535	0.0519	0.0512	0.0501	0.0488	0.0487
20	0.0549	0.0529	0.0525	0.0516	0.0499	0.0496	0.0544	0.0525	0.0520	0.0509	0.0495	0.0494
30	0.0558	0.0522	0.0517	0.0505	0.0490	0.0487	0.0545	0.0517	0.0511	0.0500	0.0484	0.0478
40	0.0560	0.0529	0.0522	0.0506	0.0494	0.0494	0.0557	0.0522	0.0514	0.0499	0.0493	0.0488
50	0.0538	0.0516	0.0510	0.0498	0.0485	0.0478	0.0528	0.0510	0.0504	0.0495	0.0470	0.0465

Given that likelihood ratio (LR) tests are optimal in the Bahadur sense [refer to Bahadur (1967); Rublík (1989)], for close alternatives from $\mathcal{G}$, the absolute local approximate Bahadur efficiency for T_n is defined as the ratio of $c_T(\theta)$ and the corresponding slope of the LR test, which equals $2K(\theta)$ – twice the Kullback–Leibler (KL) distance from the given alternative to the class of distributions within the null hypothesis [see also Meintanis et al. (2022)].

If the weak limit of T_n is $||Z||^2_{L^2}$, under certain regularity conditions [see Meintanis et al. (2022)] the local approximate Bahadur slope is equal to $c^*_T(\theta) = b''_T(0)\theta^2/(2\lambda_1) + o(\theta^2),\ \theta \to 0$. The coefficient $b_T(\theta)$ is the limit in probability of T_n/n. Therefore, the computation of λ_1 is crucial in the derivation of local approximate Bahadur efficiency.

Example 6.3 In this example, we bridge the gap in the literature by presenting Bahadur efficiencies of exponentiality test associated with kernel ρ in (13), against the following frequently considered close alternatives:

- the Weibull distribution with density

$$g_\theta(x) = e^{-x^{1+\theta}}(1+\theta)x^\theta, \quad \theta > 0, \quad x \geq 0;$$

- the gamma distribution with density

$$g_\theta(x) = \frac{x^\theta e^{-x}}{\Gamma(\theta+1)}, \quad \theta > 0 \quad x \geq 0;$$

- the Makeham distribution with density

$$g_\theta(x) = e^{-x-\theta e^x}(1+\theta e^x), \quad \theta > 0, \quad x \geq 0;$$

- the linear failure rate (LFR) distribution with density

$$g_\theta(x) = e^{-x-\theta\frac{x^2}{2}}(1+\theta x), \quad \theta > 0 \quad , x \geq 0;$$

- the mixture of exponential distributions with negative weights [EMNW (β)] with density

$$g_\theta(x;\beta) = (1+\theta)e^{-x} - \theta\beta e^{-\beta x}, \quad \theta \in \left(0, \frac{1}{\beta-1}\right], \quad x \geq 0.$$

Efficiencies of the test associated with kernel ρ are already presented in Ebner (2023), while a comprehensive comparison of exponentiality tests in terms of local approximate Bahadur efficiencies can be found in Cuparić et al. (2022b). For some recent results, see also Meintanis et al. (2022). Therefore, as a complement of previous results, in Table 19 we just present results for the test from Klar (2001). The Baha-

Table 18 Empirical levels for testing goodness-of-fit to the Poisson distribution calculated with 10,000 replications, using a parametric bootstrap (boot), $\widehat{W}_m$ (Imhof) and the Pearson System (Pearson), for several sample sizes (N)

N	Boot	Imhof		Pearson	
		$n = m = 9$	$n = 9, m = 5$	$n = m = 9$	$n = 9, m = 5$
10	0.0596	0.0537	0.0537	0.0537	0.0537
20	0.0570	0.0510	0.0510	0.0510	0.0510
30	0.0540	0.0498	0.0498	0.0498	0.0498
40	0.0511	0.0470	0.0470	0.0471	0.0471
50	0.0497	0.0492	0.0492	0.0492	0.0492

Table 19 BE of exponentially of tests from Klar (2001) in Example 4.2

		T_γ with kernel (15)			
Alternative	γ	0	1	2	3
Weibull		0.672	0.817	0.868	0.884
Gamma		0.453	0.625	0.722	0.781
Makeham		0.855	0.987	0.982	0.835
LFR		0.962	0.798	0.657	0.555
EMNW(3)		0.668	0.889	0.974	0.995

dur efficiencies are obtained using the largest eigenvalues obtained via Raylegh-Ritz method for $n = 10$.

Example 6.4 Here we reconsider tests from Example 4.3 and calculate local approximate Bahadur efficiencies against the following commonly used alternatives [see Milošević et al. (2021) and Meintanis et al. (2022)]. For the calculation of efficiencies, we use the largest eigenvalues obtained via Raylegh-Ritz method for $n = 10$. The results are presented in Table 20.

- the Lehmann alternatives

$$g_\theta^{(1)}(x) = (1+\theta)F_{\vartheta_0}^\theta(x) f_{\vartheta_0}(x)$$

- the first Ley-Paindaveine alternatives

$$g_\theta^{(2)}(x) = f_{\vartheta_0}(x) e^{-\theta(1-F_{\vartheta_0}(x))}\big(1+\theta F_{\vartheta_0}(x)\big)$$

- the second Ley-Paindaveine alternatives

$$g_\theta^{(3)}(x) = f_{\vartheta_0}(x)\big(1-\theta\pi\cos(\pi F_{\vartheta_0}(x)\big)$$

- the contamination alternatives

$$g_\theta^{(4)}(x;\mu,\sigma^2) = (1-\theta) f_{\vartheta_0}(x) + \theta\frac{1}{\sigma} f_{\vartheta_0}\Big(\frac{x-\mu}{\sigma}\Big)$$

The outcomes are displayed in Table 20. It is evident that the tests are quite sensitive to the selection of the tuning parameter, and as a result, the choice of the weight function significantly influences the behaviour of the tests. The test proposed in Dörr

Table 20 BE of normality tests in Example 4.3

	T_γ with kernel (17)				T_γ with kernel (19)			T_γ with kernel (20)			
Alternative $\quad\gamma$	1/2	1	2	3	3/2	2	3	1/2	1	2	3
$g^{(1)}$	0.637	0.809	0.918	0.951	0.386	0.543	0.831	0.916	0.964	0.959	0.964
$g^{(2)}$	0.862	0.976	0.996	0.984	0.256	0.387	0.697	0.908	0.968	0.980	0.975
$g^{(3)}$	0.955	0.986	0.939	0.899	0.194	0.299	0.387	0.802	0.875	0.895	0.883
$g^{(4)}(1,1)$	0.419	0.555	0.658	0.697	0.564	0.698	0.924	0.722	0.752	0.731	0.730
$g^{(4)}(0.5,1)$	0.552	0.719	0.838	0.880	0.449	0.604	0.883	0.862	0.905	0.895	0.900
$g^{(4)}(0,0.5)$	0.552	0.560	0.362	0.261	0.823	0.823	0.809	0.550	0.502	0.396	0.312

et al. (2021) is the least affected by the choice of tuning parameter. This test also demonstrates exceptional efficiencies against Lehmann's alternative, which are very close to the highest efficiencies of the energy and BHEP tests [refer to Meintanis et al. (2022)]. However, this test is considerably less efficient under the contamination with $N(0, 0.5)$ alternative. Conversely, in this scenario, despite being the least efficient against many alternatives, the test proposed in Henze et al. (2019) has exhibited superior performance compared to all the tests considered in the literature.

7 Comments and outlook

We highlight the high adaptability of the proposed method, which is applicable in any dimension, in discrete scenarios, and even for relatively small values of n, providing reasonable deterministic approximations. This adaptability is a distinct advantage over the competing stochastic method described in Appendix A. The Rayleigh–Ritz method can be extended to supports M that model more complex data structures, such as matrix-valued data or functional data in a Hilbert space. In these cases, it is necessary to identify an appropriate set of orthonormal basis functions with respect to the integrating measure μ. Examples of weighted L^2-type statistics in these spaces and the corresponding kernels of Gaussian processes are found in Hadjicosta and Richards (2020b) and Henze and Jiménez-Gamero (2021). Fortunately, the method remains computationally efficient when the orthonormal basis for the underlying L^2 space is implemented and performs well even for complex kernels. Although analytical solutions to the eigenvalue problem are preferable, the complexity of the formulas provided for the initial component of the kernel of the BHEP test [see Ebner and Henze (2023)] suggests that such solutions will be uncommon for more intricate kernels. Therefore, a proficient approximation, as provided by the Rayleigh–Ritz method, is sufficient for practical purposes.

Appendix A: Alternative eigenvalue approximation approaches

Appendix A1: Monte Carlo approach

This methodology was (not exclusively) introduced in Ebner and Henze (2023) and Fan et al. (2017), and is associated with the quadrature method in the traditional numerical literature, see Baker (1977), Chapter 3. It can also be found in machine learning theory, see Rasmussen and Williams (2008), and for the approximation of the spectra of Hilbert-Schmidt operators, see Koltchinskii and Giné (2000). Assume that μ is a probability measure or equivalently that the weight function w is scaled so that it is the corresponding probability density function. Then, let $Y \sim w$ be a random variable. Therefore, we can rewrite (1) as

$$\lambda f(s) = \mathbb{E}(K(s, Y) f(Y)), \quad s \in \mathbb{R}. \tag{29}$$

An empirical counterpart to (29) is found by letting $y_1, y_2, \dots, y_N$, $N \in \mathbb{N}$, be independent realizations of Y and approximating the expected value in (29) by

$$\mathbb{E}(K(s,Y)f(Y)) \approx \frac{1}{N}\sum_{j=1}^{N} K(s,y_j)f(y_j), \quad s \in \mathbb{R}. \tag{30}$$

Evaluate (30) at the points $y_1, \dots, y_n$ to get

$$\lambda f(y_i) = \frac{1}{N}\sum_{j=1}^{N} K(y_i,y_j)f(y_j), \quad i = 1, \dots, N, \tag{31}$$

which is a system of N linear equations. Writing $v = (f(y_1), \dots, f(y_N)) \in \mathbb{R}^N$ and $\widetilde{K} = (K(y_i,y_j)/N)_{i,j=1,\dots,N} \in \mathbb{R}^{N\times N}$, we can rewrite (31) in matrix form

$$\widetilde{K}v = \lambda v \tag{32}$$

from which the (approximated) eigenvalues $\lambda_1, \dots, \lambda_N$ can be computed explicitly. Note that for every eigenvalue λ_j we have an eigenvector (say) $v_j \in \mathbb{R}^N$, whose components are the (approximated) values of the eigenfunctions (say) f_j evaluated at $y_1, \dots, y_N$. Obviously $\widetilde{K}$ is a random matrix, so that the calculated eigenvalues are random as well.

Appendix A2: Matrix-based operator' approximation approach

In Božin et al. (2020) the authors proposed the method for approximation of eigenvalues which is essentially based on the following steps: Recall (1), namely

$$\lambda f(s) = \int_M K(s,t)f(t)\mu(\mathrm{d}t), \quad s \in M,$$

with $\mu(\mathrm{~d~}t) = w(t)\mathrm{d}t$ and $\int_M tw(t)\mathrm{d}t < \infty$.

1. Consider the symmetrized operator

$$\lambda f(s) = \int_M K(s,t)f(t)\sqrt{w(t)w(s)}\mathrm{d}t, \quad s \in M, \tag{33}$$

which possesses the same spectrum as the one defined by (1).

2. If $M = [-A, A]$, consider the sequence of symmetric linear operators defined by $(2m+1)\times(2m+1)$ matrices $M_\omega^{(m)} = ||m_{i,j}^{(m)}||,\ |i| \le m, |j| \le m$ where

$$m_{i,j}^{(m)} = K\left(\frac{iA}{m}, \frac{jA}{m}\right)\sqrt{w\left(\frac{iA}{m}\right)w\left(\frac{jA}{m}\right)}.$$

This sequence of operators converges in norm to the operator defined by (33), and hence the spectra of these two operators are at a distance that tends to zero.Observe that if $M = [a, b]$ for arbitrary $a < b$, the operator is suitably adjusted. In the case of unbounded support, step 2 is performed after truncating the operator (33) in such a manner that the resulting operator differs from the initial one on a set of negligible measure.

While this method is effective and straightforward to understand, one of its main disadvantages is that it may not demonstrate optimal computational efficiency, especially when dealing with complex kernel functions.

We demonstrate how this method operates using the kernels from Examples 4.2 and 4.3. The eigenvalues associated with the tests from 4.2 and 4.3 are displayed in Tables 21 and 22 respectively.

Table 21 Approximation of the largest eigenvalue $\lambda_1(\gamma)$ from Example 4.2 with $A = 10$

Kernel	γ	$m = 100$	$m = 500$	$m = 1000$	$m = 2000$	$m = 5000$
(13)	0	2.70032e−1	2.71923e−1	2.72186e−1	2.7232e−1	2.72401e−1
	1	1.00554e−1	1.01133e−1	1.01227e−1	1.01276e−1	1.01306e−1
	2	5.24446e−2	5.26588e−2	5.27049e−2	5.27296e−2	5.27449e−2
	3	3.22069e−2	3.22729e−2	3.22988e−2	3.23134e−2	3.23227e−2
(15)	0	1.93010e−1	1.94551e−1	1.94746e−1	1.94843e−1	1.94901e−1
	1	3.06451e−2	3.08897e−2	3.09206e−2	3.09361e−2	3.09453e−2
	2	8.82674e−3	8.89719e−3	8.90608e−3	8.91053e−3	8.91320e−3
	3	3.43041e−3	3.45777e−3	3.46122e−3	3.46295e−3	3.46399e−3

Table 22 Approximation of the largest eigenvalue $\lambda_1(\gamma)$ from Example 4.3

Kernel	γ	A	$m = 100$	$m = 500$	$m = 1000$	$m = 2000$	$m = 5000$
(17)	0.5	5	1.19722	1.20201	1.20261	1.20291	1.20309
	1	4	4.49887e−1	4.51685e−1	4.51910e−1	4.52023e−1	4.52091e−1
	2	3	1.40959e−1	1.41522e−1	1.41593e−1	1.41628e−1	1.41649e−1
	3	3	6.55722e−2	6.58342e−2	6.58671e−2	6.58836e−2	6.58934e−2
(19)	1.5	4	3.36331e−1	3.37675e−1	3.37843e−1	3.37928e−1	3.37979e−1
	2	3	8.72568e−2	8.76048e−2	8.76485e−2	8.76704e−2	8.76835e−2
	3	3	1.38116e−2	1.38668e−2	1.38738e−2	1.38772e−2	1.38793e−2
(20)	0.5	5	9.62013e−1	9.65857e−1	9.66340e−1	9.66582e−1	9.66728e−1
	1	4	5.97812e−1	6.00201e−1	6.00501e−1	6.00651e−1	6.00741e−1
	2	3	3.96036e−1	3.97618e−1	3.97817e−1	3.97916e−1	3.97976e−1
	3	3	3.08925e−1	3.10160e−1	3.10315e−1	3.10392e−1	3.10439e−1

Appendix B: Local approximate Bahadur efficiency

In order to compute the approximate Bahadur efficiency, we need the following conditions to be met:

1. T_n converges in distribution to a non-degenerate distribution function F,
2. $\log(1-F(t)) = -\frac{a_T t^2}{2}(1+o(1)),\ \to\infty,$
3. The limit in probability under an alternative from $\mathcal{G}$ $\lim_{n\to\infty} T_n/\sqrt{n} = b_T(\theta) > 0$ exists for $\theta \in \Theta_1$.Then, we have

$$c_T^*(\theta) = a_T b_T^2(\theta) \tag{34}$$

which represents the approximate Bahadur slope of T_n. Typically, under a close alternative where certain regularity conditions are met, i.e., when $\theta \to 0$, $b_T(\theta)$ can be expanded into a Maclaurin series. In the case of test statistics with an asymptotic normal distribution, the coefficient a_T is equal to the inverse of the asymptotic variance. However, when the distribution is that of $\sum_{j=1}^{\infty} \lambda_j N_j^2,\ \lambda_1 \geq \lambda_2 \geq \cdot$, the coefficient a_T, which corresponds to $\sqrt{T_n}$, is equal to λ_1^{-1} [for a detailed explanation, see Cuparić et al. (2022b)].

It's important to note that the local approximate Bahadur efficiency is an approximation of the local exact Bahadur efficiency, a measure that is significantly more complex to calculate in practice. While Yu and Nikitin and I. Peaucelle. (2004) provided results for test statistics in the form of U- or V-statistics, the case involving unknown parameters still remains an open question.

Acknowledgements The authors thank the editor and two anonymous referees for their constructive comments and suggestions which helped to improve the presentation. The authors thank Norbert Henze for valuable comments that lead to a better readability of the paper. The research of B. Ebner and B. Milošević was supported by the bilateral cooperation project "Modeling complex data - Selection and Specification" between the Federal Republic of Germany and the Republic of Serbia (the contract 337-00-19/2023-01/6). The project underlying this article was funded by the Federal Ministry of Education and Research of Germany, the Ministry of Science, Technological Development and Innovation of Republic of Serbia (the contract 451-03-136/2025-03/200104), and COST action CA21163 - Text, functional and other high-dimensional data in econometrics: New models, methods, applications (HiTEc). The research of M.D. Jiménez-Gamero was supported by grant PID2023-148811NB-I00, funded by MICIU/AEI/10.13039/501100011033 and ERDF/EU.

Funding Open Access funding enabled and organized by Projekt DEAL.

References

Aleksić DG, Milošević B (2025) To impute or not? Testing multivariate normality on incomplete dataset: revisiting the BHEP test. J Appl Stat 52(9):1742–1759

Allison J, Milošević B, Obradović M, Smuts M (2022) Distribution-free goodness-of-fit tests for the Pareto distribution based on a characterization. Comput Stat 37(1):403–418

Allison JS, Ebner B, Smuts M (2023) Logistic or not logistic? Stat Neerl 77(4):429–443

Allison JS, Pretorius C (2017) A Monte Carlo evaluation of the performance of two new tests for symmetry. Comput Stat 32(4):1323–1338

Bahadur R (1967) An optimal property of the likelihood ratio statistic. In Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability 1:13–26

Bahraoui T, Bouezmarni T, Quessy J-F (2018) A family of goodness-of-fit tests for copulas based on characteristic functions. Scand J Stat 45(2):301–323

Baker CTH (1977) The numerical treatment of integral equations. Oxford University Press, Oxford

Baringhaus L (1996) Fibonacci numbers, Lucas numbers and integrals of certain Gaussian processes. Proc Am Math Soc 124(12):3875–3884

Baringhaus L, Ebner B, Henze N (2017) The limit distribution of weighted L^2-goodness-of-fit statistics under fixed alternatives, with applications. Ann Inst Stat Math 69(5):969–995

Baringhaus L, Gaigall D, Thiele JP (2018) Statistical inference for L^2-distances to uniformity. Comput Stat 33(4):1863–1896

Baringhaus L, Henze N (2008) A new weighted integral goodness-of-fit statistic for exponentiality. Stat Probab Lett 78(8):1006–1016

Becker M, Klüßner S (2024) PearsonDS: Pearson Distribution System. R package version 1.3.1

Betsch S, Ebner B (2019) A new characterization of the gamma distribution and associated goodness-of-fit tests. Metrika 82(7):779–806

Bilodeau M, de Micheaux PL (2005) A multivariate empirical characteristic function test of independence with normal marginals. J Multivar Anal 95(2):345–369

Bose A, Kappara D, Bhattacharjee M (2023) Estimation of Bergsma's covariance. J Korean Stat Soc 52(4):1025–1054

Božin V, Milošević B, Nikitin YY, Obradović M (2020) New characterization-based symmetry tests. Bull Malaysian Math Sci Soc 43(1):297–320

Cuparić M, Milošević B, Obradović M (2022) Asymptotic distribution of certain degenerate V-and U-statistics with estimated parameters. Math Commun 27(1):77–100

Cuparić M, Milošević B, Obradović M (2022) New consistent exponentiality tests based on V-empirical Laplace transforms with comparison of efficiencies. Revista de la Real Academia de Ciencias Exactas Físicas y Naturales Serie A Matemáticas 116(1):1–26

Deheuvels P, Martynov G (2003) Karhunen-Loève expansions for weighted Wiener processes and Brownian bridges via Bessel functions. In: Hoffmann-Jørgensen J, Wellner JA, Marcus MB (eds) High dimensional probability III. pp. Birkhäuser Basel, Basel, pp 57–93

Duan J, Wang W (2014) Effective dynamics of stochastic partial differential equations. Academic Press, Boston

Duchesne P, de Micheaux PL (2010) Computing the distribution of quadratic forms: further comparisons between the Liu-Tang-Zhang approximation and exact methods. Comput Stat Data Anal 54:858–862

Dörr P, Ebner B, Henze N (2021) Testing multivariate normality by zeros of the harmonic oscillator in characteristic function spaces. Scand J Stat 48(2):456–501

Ebner B (2020) On combining the zero bias transform and the empirical characteristic function to test normality. ALEA 18:1029–1045

Ebner B (2023) The test of exponentiality based on the mean residual life function revisited. J Nonparametric Stat 35(3):601–621

Ebner B, Henze N (2023) Bahadur efficiencies of the Epps-Pulley test for normality. J Math Sci 273(5):861–870

Ebner B, Henze N (2023) On the eigenvalues associated with the limit null distribution of the Epps-Pulley test of normality. Stat Pap 64(3):739–752

Ebner B, Liebenberg SC, Visagie IJH (2022) A new omnibus test of fit based on a characterization of the uniform distribution. Statistics 56(6):1364–1384

Fan Y, Lafaye de Micheaux P, Penev S, Salopek D (2017) Multivariate nonparametric test of independence. J Multivar Anal 153:189–210

Fasshauer GE, McCourt M (2016) Kernel-based approximation methods using MATLAB. Interdisciplinary mathematical sciences; 19. World Scientific, New Jersey
Gautschi W (2004) Orthogonal polynomials: computation and approximation. Numerical mathematics and scientific computation. Oxford University Press, Oxford
Gürtler N, Henze N (2000) Recent and classical goodness-of-fit tests for the Poisson distribution. J Stat Pn Inference 90(2):207–225
Hadjicosta E, Richards D (2020) Integral transform methods in goodness-of-fit testing, I: the gamma distributions. Metrika 83(7):733–777
Hadjicosta E, Richards D (2020) Integral transform methods in goodness-of-fit testing, II: the Wishart distributions. Ann Inst Stat Math 72(6):1317–1370
Henze N (1990) An approximation to the limit distribution of the Epps-Pulley test statistic for normality. Metrika 37(1):7–18
Henze N, Jiménez-Gamero MD (2019) A new class of tests for multinormality with i.i.d. and GARCH data based on the empirical moment generating function. TEST 28(2):499–521
Henze N, Jiménez-Gamero MD (2021) A test for Gaussianity in Hilbert spaces via the empirical characteristic functional. Scand J Stat 48(2):406–428
Henze N, Jiménez-Gamero MD, Meintanis SG (2019) Characterizations of multinormality and corresponding tests of fit, including for GARCH models. Economet Theor 35(3):510–546
Henze N, Mayer C (2020) More good news on the HKM test for multivariate reflected symmetry about an unknown centre. Ann Inst Stat Math 72:741–770
Henze N, Nikitin YY (2000) A new approach to goodness-of-fit testing based on the integrated empirical process. J Nonparametric Stat 12(3):391–416
Henze N, Wagner T (1997) A new approach to the BHEP tests for multivariate normality. J Multivar Anal 62(1):1–23
Horváth L, Kokoszka P (2012) Inference for functional data with applications, vol 200. Springer, Cham
Hušková M, Neumeyer N, Niebuhr T, Selk L (2019) Specification testing in nonparametric AR-ARCH models. Scand J Stat 46(1):26–58
Imhof JP (1961) Computing the distribution of quadratic forms in normal variables. Biometrika 48:419–426
Jammalamadaka SR, Jiménez-Gamero MD, Meintanis SG (2019) A class of goodness-of-fit tests for circular distributions based on trigonometric moments. SORT 43(2):317–336
Jiménez-Gamero M, Alba-Fernández M, Jodrá P, Barranco-Chamorro I (2015) An approximation to the null distribution of a class of Cramér–von Mises statistics. Math Comput Simul 118:258–272
Kanwal RP (1971) Linear integral equations: theory and technique. Academic Press, New York
Klar B (2001) Goodness-of-fit tests for the exponential and the normal distribution based on the integrated distribution function. Ann Inst Stat Math 53(2):338–353
Koltchinskii V, Giné E (2000) Random matrix approximation of spectra of integral operators. Bernoulli 6(1):113–167
Mardia K, Jupp P (2000) Directional statistics. Wiley series in probability and statistics. John Wiley & Sons Ltd, Chichester
Matsui M, Takemura A (2005) Empirical characteristic function approach to goodness-of-fit tests for the Cauchy distribution with parameters estimated by MLE or EISE. Ann Inst Stat Math 57(1):183–199
Meintanis S, Milošević B, Obradović M (2022) Bahadur efficiency for certain goodness-of-fit tests based on the empirical characteristic function. Metrika 86(7):723–751
Milošević B (2020) Asymptotic efficiency of goodness-of-fit tests based on Too-Lin characterization. Commun Stat Simul Comput 49(8):2082–2101
Milošević B, Nikitin YY, Obradović M (2021) Bahadur efficiency of EDF based normality tests when parameters are estimated. Zapiski Nauchnykh Seminarov POMI 501:203–217
Nikitin YY (1995) Asymptotic efficiency of nonparametric tests. Cambridge University Press, New York
Yu Y, Nikitin, Peaucelle I (2004) Efficiency and local optimality of nonparametric tests based on U- and V-statistics. Metron 62(2):185–200
Novoa-Muñoz F, Jiménez-Gamero MD (2016) A goodness-of-fit test for the multivariate Poisson distribution. SORT 40(1):113–138
Olver FWJ, Lozier DW, Boisvert RF, Clark CW, editors. (2010) NIST handbook of mathematical functions. U.S. Department of Commerce, National Institute of Standards and Technology, Washington, DC; Cambridge University Press, Cambridge
Rasmussen CE, Williams CKI (2008) Gaussian processes for machine learning. Adaptive computation and machine learning. MIT Press, Cambridge, MA

Rivas Martínez GI, Jiménez-Gamero MD (2021) Computationally efficient approximations for independence tests in non-parametric regression. J Stat Comput Simul 91(6):1134–1154
Rivas-Martínez GI, Jiménez-Gamero MD, Moreno-Rebollo JL (2019) A two-sample test for the error distribution in nonparametric regression based on the characteristic function. Stat Pap 60(4):1369–1395
Rublík F (1989) On optimality of the LR tests in the sense of exact slopes. I General Case Kybernetika 25(1):13–14
Rueda R, O'Reilly F (1999) Tests of fit for discrete distributions based on the probability generating function. Comm Statist Simulation Comput 28(1):259–274
Rueda R, O'Reilly F, Pérez-Abreu V (1991) Goodness of fit for the Poisson distribution based on the probability generating function. Commun Stat Theory Methods 20(10):3093–3110
Shi H, Drton M, Han F (2022) Distribution-free consistent independence tests via center-outward ranks and signs. J Am Stat Assoc 117(537):395–410
Shorack GR, Wellner JA (1986) Empirical processes with applications to statistics. Wiley, New York
Stephens MA (1970) Use of the Kolmogorov-Smirnov, Cramér-von Mises and related statistics without extensive tables. J Roy Stat Soc: Ser B (Methodol) 32(1):115–122 (**12**)
Stephens MA (1976) Asymptotic results for goodness-of-fit statistics with unknown parameters. Ann Stat 4(2):357–369
Stephens MA (1977) Goodness of fit for the extreme value distribution. Biometrika 64(3):583–588
Williams C, Shawe-Taylor JS (2003) The stability of kernel principal components analysis and its relation to the process eigenspectrum. In: Becker S, Thrun S, Obermayer K (eds) Advances in Neural Information Processing Systems 15. MIT Press, Cambridge, pp 383–390