



Computing the asymptotic distribution of second-order U - and V -statistics

Raffaello Seri ^{a,b,*}

^a Center for Nonlinear and Complex Systems, Università degli Studi dell'Insubria, Via Valleggio 11, 22100 Como, Italy

^b DiECO, Università degli Studi dell'Insubria, Via Monte Generoso 71, 21100 Varese, Italy



ARTICLE INFO

Article history:

Received 28 February 2021

Received in revised form 19 January 2022

Accepted 20 January 2022

Available online 14 February 2022

Keywords:

Wielandt–Nyström method

Clenshaw–Curtis quadrature

U - and V -statistics

ABSTRACT

Under general conditions, the asymptotic distribution of degenerate second-order U - and V -statistics is an (infinite) weighted sum of χ^2 random variables whose weights are the eigenvalues of an integral operator associated with the kernel of the statistic. Also the behavior of the statistic in terms of power can be characterized through the eigenvalues and the eigenfunctions of the same integral operator. No general algorithm seems to be available to compute these quantities starting from the kernel of the statistic. An algorithm is proposed to approximate (as precisely as needed) the asymptotic distribution and to build several measures of performance for tests based on U - and V -statistics. The algorithm uses the Wielandt–Nyström method of approximation of an integral operator based on quadrature, and can be used with several methods of numerical integration. An extensive numerical study shows that the Wielandt–Nyström method based on Clenshaw–Curtis quadrature performs very well both for the eigenvalues and the eigenfunctions.

© 2022 Elsevier B.V. All rights reserved.

1. Introduction

Distributions given by weighted sums of χ^2 random variables arise in many situations such as χ^2 tests with estimated parameters (see van der Vaart, 1998, p. 249) or non-standard normalizations (see Colombi, 2020), the likelihood ratio and variance statistics in Vuong (1989), the test for the rank of a matrix in Robin and Smith (2000) and the asymptotic distribution of the entropy in Seri and Martinoli (2021). The most prominent one, however, is the asymptotic distribution of degenerate second-order U - and V -statistics that is, in general, an (infinite) weighted sum of χ^2 random variables. The eigenvalues and the eigenfunctions of an integral operator play a central role in the asymptotic distribution and power of the statistics. An algorithm, preliminarily studied in Seri and Choirat (2012), is provided for the numerical approximation of the eigenvalues and of the eigenfunctions of this operator, and of the cumulative distribution function (cdf) of the asymptotic distribution. The algorithm can also be used to approximate (as precisely as needed) the asymptotic power of the test statistics and to build several measures of performance for tests based on U - and V -statistics. Since the explicit identification of the eigenfunctions is a very difficult task, these methods can be used to approximate the eigenfunctions and to identify them with known functions; after these steps, verifying that they respect the integral equation is usually much simpler.

The results cover the one-dimensional case. Moreover, only kernels defined on a bounded interval (that will in most cases be reduced to $[0, 1]$) will be considered. The more general case of infinite intervals can be obtained through the application of the probability integral transform.

* Correspondence to: DiECO, Università degli Studi dell'Insubria, Via Monte Generoso 71, 21100 Varese, Italy.

E-mail address: raffaello.seri@uninsubria.it.

Special cases of the algorithm have already been proposed in the statistical literature. A first instance is in Schilling (1983b, p. 23), who proposes a version of this algorithm in the case of a nearest-neighbor goodness-of-fit test and evaluates its accuracy by comparing the theoretical moments of the asymptotic distribution with those of the resulting approximated statistic. In Csörgő (1986, p. 719), the author suggests the use of the same algorithm of Schilling (1983b) for another test statistic. In Koltchinskii and Giné (2000), the authors propose to approximate the spectrum of an integral operator using the evaluations of the kernel on a random sample of independent points. In Choirat and Seri (2006), random points are replaced by quasi-Monte Carlo points, while in Adamczak and Bednorz (2015) they are replaced by realizations of a Markov chain.

The methods proposed in these papers consist in approximating the integral appearing in the integral operator through a sum over a finite set of points. However, the performance of this method can be greatly improved if the contribution of the points to the sum is allowed to be multiplied by a weight, that is if the sum is replaced by a weighted sum. The procedure is often used in Numerical Analysis under the name of Wielandt–Nyström method for the approximation of the eigenvalues and eigenfunctions of an integral operator. In this context the technique is often applied using Gauss–Legendre quadrature: in Delves and Mohamed (1985, pp. 245–246), it is stated that “such results are enough to make a numerical analyst weep.” One of the aims of the present paper is to show that this would not be the case for a statistician since Gauss–Legendre quadrature is not the best possible choice.

In particular, several methods of integration (namely Monte Carlo integration, quasi-Monte Carlo integration through the Halton or van der Corput sequence, the trapezium rule, and Gauss–Legendre and Clenshaw–Curtis quadrature rules) are compared. Monte Carlo integration performs very poorly as expected by the bounds in Koltchinskii and Giné (2000, Section 4). Quasi-Monte Carlo integration performs better, especially as concerns eigenvalues but shows some strange behavior when it comes to the computation of eigenfunctions. But the most interesting result concerns the comparison of Gauss–Legendre and Clenshaw–Curtis quadrature rules: it is usually stated in the literature that Gauss–Legendre integration has a factor-of-2 advantage for finite N , since the $(N + 1)$ -point Gauss–Legendre quadrature rule integrates exactly polynomials of degree $2N + 1$ while the $(N + 1)$ -point Clenshaw–Curtis quadrature rule only integrates exactly polynomials of degree N . However, this difference is rarely observed in practice: see Trefethen (2008) and Xiang and Bornemann (2012) for historical reviews of the literature about this fact and theoretical explanations. On the other hand, Clenshaw–Curtis quadrature has an edge on Gauss–Legendre quadrature from a computational point of view since the N -point Clenshaw–Curtis formula can be implemented in $O(N \ln N)$ operations through the Fast Fourier Transform, while the N -point Gauss–Legendre formula requires $O(N^2)$ operations by solving a tridiagonal eigenvalue problem, or $O(N^3)$ when the tridiagonal structure of the eigenvalue problem is neglected; see Trefethen (2008) for more details on this point.

Therefore, the computational study provides a further case in which Gauss–Legendre and Clenshaw–Curtis quadrature rules behave in very similar ways. Some heuristic reasonings about theoretical convergence rates but no formal proof are provided, since the problem is often complicated by the presence of a discontinuity along the diagonal of the kernel (so that the kernel does not even belong to \mathcal{C}^1).

Section 2 briefly reviews some asymptotic results for degenerate U - and V -statistics involving quantities that can be computed using the algorithm described in Section 3. The convergence properties of the algorithm are evaluated both theoretically, in Section 4, and empirically, in Section 5, through the systematic application to three well known V -statistics, namely the Watson, Cramér–von Mises and Anderson–Darling test statistics. In the same section, the method is applied to a statistic proposed in Hall (1985) and to a class of statistics described in Bickel and Breiman (1983) and Schilling (1983a,b). Section 6 concludes the paper. A description of the algorithm for the computation of the cdf and the proofs of the results of the paper are contained in Appendix A.

2. Asymptotic results for degenerate U - and V -statistics

Let (X_1, \dots, X_n) be an independent and identically distributed (iid) sample from the probability \mathbb{P} defined on the measurable space $(\mathbf{X}, \mathcal{B}(\mathbf{X}))$. The V - or von Mises statistic is defined as:

$$V_n = \frac{1}{n^2} \sum_{k, \ell=1}^n h(x_k, x_\ell)$$

and the U -statistic as:

$$U_n = \frac{1}{n(n-1)} \sum_{k \neq \ell=1}^n h(x_k, x_\ell)$$

where $h : \mathbf{X} \times \mathbf{X} \rightarrow \mathbb{R}$ (or, more rarely, \mathbb{C}) is called the *kernel* of the statistic. There is no loss of generality in taking h symmetric, i.e. $h(x_k, x_\ell) = h(x_\ell, x_k)$, as otherwise one could set $\tilde{h}(x_k, x_\ell) = \frac{h(x_k, x_\ell) + h(x_\ell, x_k)}{2}$. The name of *quadratic statistics* (see, e.g., Gregory, 1980) is sometimes used to indicate both U - and V -statistics. The present results do not cover higher-order statistics such as, e.g., $V_n = \frac{1}{n^3} \sum_{k, \ell, m=1}^n h(x_k, x_\ell, x_m)$.

The asymptotic behavior of these statistics is different if h respects the following condition, called *degeneracy* and assumed throughout the rest of the paper, or if it does not:

$$(H) \quad \mathbb{E}_{(\ell)} h(x_k, X_\ell) = \mathbb{E}_{(k,\ell)} h(X_k, X_\ell) = 0, \text{ for } k \neq \ell \text{ and } \mathbb{P}\text{-almost any } x_k.$$

The notation $\mathbb{E}_{(\ell)} h(x_k, X_\ell)$ denotes that the expectation is taken with respect to the distribution of X_ℓ . The condition $\mathbb{E}_{(k,\ell)} h(X_k, X_\ell) = 0$ is often verified for the statistics considered here, but it can otherwise be obtained by centering.

The asymptotic distribution for degenerate U - and V -statistics has been worked out by von Mises (1947), Filippova (1962), Gregory (1977), Serfling (1980) and in greater generality by Rubin and Vitale (1980), Denker (1982), Chen and White (1998), Dewan and Prakasa Rao (2001, 2002). In order to characterize the asymptotic distribution, some properties of the integral operator \mathcal{H} defined by:

$$\mathcal{H}\phi(x) = \int h(x, y)\phi(y)\mathbb{P}(dy) \quad (1)$$

are needed. The kernel is said to be *Hermitian* if $h(x, y) = \overline{h(y, x)}$. The operator \mathcal{H} is said to be *Hilbert–Schmidt* if $\iint |h(x, y)|^2 \mathbb{P}(dx)\mathbb{P}(dy) < \infty$, and *self-adjoint* if it is Hilbert–Schmidt and if the kernel is Hermitian.

If the operator is self-adjoint, from Exercise 44 on p. 1083 and Exercise 56 on p. 1087 in Dunford and Schwartz (1963), the following spectral decomposition of the kernel h holds:

$$h(x, y) = \sum_{j=1}^{\infty} \lambda_j \phi_j(x) \overline{\phi_j(y)} \quad (2)$$

where $(\lambda_j)_j$ are the real eigenvalues and $(\phi_j)_j$ are the orthonormal eigenfunctions of the operator \mathcal{H} . In most cases, the kernel assumes real values, is *symmetric*, i.e. $h(x, y) = h(y, x)$, and *positive semidefinite*, i.e. $\sum_{k,\ell=1}^N h(y_k, y_\ell) w_k w_\ell \geq 0$ for all finite sequences of points $(y_k)_{k=1,\dots,N}$ and weights $(w_k)_{k=1,\dots,N}$. In this case, the eigenvalues are non-negative, i.e. $\lambda_j \geq 0$ for any j . However, when possible, the results will be stated for the more general case in which the kernel is not necessarily positive semidefinite.

The distribution can be obtained under degeneracy and mild additional assumptions, namely $\mathbb{E}|h(X_1, X_2)|^2 < \infty$ for U -statistics, and $\mathbb{E}|h(X_1, X_2)|^2 < \infty$ and $\mathbb{E}|h(X_1, X_1)| < \infty$ for V -statistics. It is possible to show that U - and V -statistics have asymptotic distributions that are weighted (infinite) sums of χ^2 random variables:

$$\begin{aligned} nU_n &\xrightarrow{\mathcal{D}} \sum_{j=1}^{\infty} \lambda_j (Z_j^2 - 1), \\ nV_n &\xrightarrow{\mathcal{D}} \sum_{j=1}^{\infty} \lambda_j Z_j^2, \end{aligned} \quad (3)$$

where $(Z_j)_j$ are independent standard normal random variables (see, e.g., Theorem on p. 194 in Serfling, 1980, and Theorem 12.10 on p. 169 in van der Vaart, 1998, for U -statistics, and Theorem 2.3 in Gregory, 1977, with $h_n \equiv 0$, for V -statistics).

As an example, consider the Cramér–von Mises statistic for the hypothesis that a sample is extracted from a distribution with continuous cdf F on \mathbb{R} . Let F_n be the empirical cdf based on the sample (X_1, \dots, X_n) . The Cramér–von Mises statistic is defined as:

$$\begin{aligned} V_n &= \int_{-\infty}^{+\infty} (F_n(x) - F(x))^2 F(dx) \\ &= \frac{1}{n^2} \sum_{i,j=1}^n \left(\frac{F^2(X_i)}{2} + \frac{F^2(X_j)}{2} - F(X_i) \vee F(X_j) + \frac{1}{3} \right). \end{aligned}$$

Under the null hypothesis, the sample $(F(X_1), \dots, F(X_n))$ is independently and uniformly distributed and, therefore, the distribution can be reduced to the one of:

$$V_n = \frac{1}{n^2} \sum_{i,j=1}^n \left(\frac{U_i^2}{2} + \frac{U_j^2}{2} - U_i \vee U_j + \frac{1}{3} \right)$$

where (U_1, \dots, U_n) is an independently and uniformly distributed sample. The asymptotic distribution (see, e.g., Serfling, 1980, p. 64 or van der Vaart, 1998, pp. 170–171) is:

$$nV_n \xrightarrow{\mathcal{D}} \sum_{j=1}^{\infty} \frac{Z_j^2}{j^2 \pi^2}.$$

Now, the asymptotic distribution of V_n under a Pitman drift (see Gregory, 1977, Theorems 2.1 and 2.3) will be considered. The objective is to study the behavior of the test statistic under the probability measure \mathbb{P}_n^* where $\frac{d\mathbb{P}_n^*}{d\mathbb{P}} = 1 + n^{-\frac{1}{2}} g_n$, for

a sequence $(g_n)_n \in L^2(\Omega, \mathcal{A}, \mathbb{P})$ converging to $g \in L^2(\Omega, \mathcal{A}, \mathbb{P})$ and \mathbb{P} the probability measure such that the kernel h is degenerate under it. Then, if $\mathbb{E}|h(X_1, X_2)|^2 < \infty$ and $\mathbb{E}|h(X_1, X_1)| < \infty$:

$$nV_n \xrightarrow{\mathcal{D}} \sum_{j=1}^{\infty} \lambda_j (Z_j + a_j)^2,$$

where $(Z_j)_j$ are independent standard normal random variables and $a_j = \int g\phi_j d\mathbb{P}$, for $j \geq 1$. This shows that the eigenfunctions of \mathcal{H} play a very important role in the determination of the power of quadratic statistics. It is easy to see, using orthonormality of the $(\phi_j)_j$, that the coefficients $(a_j)_j$ are the same as in the expansion $g = \sum_{j=1}^{\infty} a_j \phi_j$.

In Gregory (1977, p. 120), a simple index of performance for a test based on U - and V -statistics is introduced:

$$e_h = \frac{\sum \lambda_j a_j^2}{\left(2 \sum \lambda_j^2\right)^{\frac{1}{2}}}.$$

In general, the ratio of these indexes for two tests is the true asymptotic relative efficiency, that is the limit of the ratio of sample sizes that gives the same limit of the power for the two tests. Other measures of efficiency are discussed in Gregory (1980, pp. 121 ff.); in particular, it is worthwhile to remark condition (3.12) (b) on p. 126, in which optimality of a quadratic test with respect to Bahadur efficiency is shown to be equivalent to the fact that the eigenfunction corresponding to the largest eigenvalue is proportional to a certain function h_0 .

The eigenfunctions help characterize the behavior of the statistic also under a fixed alternative of the form $\frac{d\mathbb{P}^*}{d\mathbb{P}} = 1 + \sum_{j=1}^{\infty} a_j \phi_j$, as shown in Hall (1985, p. 129) (note that U_n and V_n are non-degenerate quadratic statistics with respect to \mathbb{P}^*). In this case, if the eigenvalues $(\lambda_j)_j$ are positive and $\lambda_1 \geq \lambda_j > 0$ for any j , it can be shown that $\frac{\sum \lambda_j a_j^2}{\lambda_1}$ is a measure of efficiency of quadratic statistics.

The role of the eigenfunctions in determining power is the basis for the theory of the components of quadratic statistics started, for the Cramér-von Mises statistic, in Durbin and Knott (1972) and Durbin et al. (1975, 1977). The authors express this statistic as a functional of an empirical process and not as a V -statistic, but their reasoning can be extended to a general V -statistic. If the eigenvalues $(\lambda_j)_j$ are non-negative and the kernel h of a V -statistic is replaced with its spectral decomposition (2), nV_n can be written as:

$$\begin{aligned} nV_n &= \frac{1}{n} \sum_{k, \ell=1}^n h(x_k, x_\ell) = \frac{1}{n} \sum_{k, \ell=1}^n \sum_{j=1}^{\infty} \lambda_j \phi_j(x_k) \phi_j(x_\ell) \\ &= \frac{1}{n} \sum_{j=1}^{\infty} \lambda_j \sum_{k=1}^n \phi_j(x_k) \sum_{\ell=1}^n \phi_j(x_\ell) \\ &= \sum_{j=1}^{\infty} \left[\frac{1}{\sqrt{n}} \sum_{k=1}^n \sqrt{\lambda_j} \phi_j(x_k) \right]^2. \end{aligned}$$

Any term $\frac{1}{\sqrt{n}} \sum_{k=1}^n \sqrt{\lambda_j} \phi_j(x_k)$ is a component of the V -statistic. Each component accounts for a contribution to the distance between the sample and the null hypothesis and a contribution to the power of the statistic against an alternative hypothesis. Schoenfeld (1977) develops a theory of asymptotically most powerful tests against a sequence of alternative hypotheses converging to the null. These tests are given by linear combinations of the previously quoted components and are therefore asymptotically normal. Parr and Schucany (1982) establish a link between efficiency and robustness on one side and components of minimum distance objective functions on the other. Also the results about asymptotic power of tests based on U -statistics in Hall (1985) exploit the decomposition of the kernel in eigenvalues and eigenfunctions and of the statistic in components. The eigenfunctions can also be used to derive smooth goodness-of-fit tests as in Neyman (1937), see also Durbin and Knott (1972, p. 302).

The previous discussion points at the interest of eigenvalues and eigenfunctions of \mathcal{H} in establishing the asymptotic distribution and power of a quadratic statistic. The methods presented here can be used to derive the components of a generic quadratic statistic and to extend the analysis beyond the cases in which the eigenfunctions are explicitly known.

3. The algorithm

In this section the algorithm for the approximation of the eigenvalues and the eigenfunctions of the integral operator is described.

First, the integral operator \mathcal{H} of equation (1) is replaced by an approximate one (say \mathcal{H}_N) yielding N eigenvalues $(\hat{\lambda}_{N,j})_{j=1, \dots, N}$: this induces two sources of errors, namely the neglection of the smallest eigenvalues (i.e. $(\lambda_j)_{j=N+1, \dots}$)

and the numerical approximation of the leading N ones. This step also provides approximations to the eigenfunctions of the integral operator. This is explained in Section 3.1.

Second, if the objective is to derive the cdf, the distribution of $\sum_{j=1}^{\infty} \lambda_j Z_j^2$ can be replaced by the distribution of $\sum_{j=1}^N \hat{\lambda}_{N,j} Z_j^2$, where $(Z_j)_j$ are independent standard normal random variables. If the weights $(\lambda_j)_j$ and $(\hat{\lambda}_{N,j})_j$ are non-negative, it can be shown (see Appendix A.1) that:

$$\sup_{x \geq 0} \left| \mathbb{P} \left(\sum_{j=1}^N \hat{\lambda}_{N,j} Z_j^2 \leq x \right) - \mathbb{P} \left(\sum_{j=1}^{\infty} \lambda_j Z_j^2 \leq x \right) \right| = O \left(\sum_{j=1}^{\infty} |\hat{\lambda}_{N,j} - \lambda_j| \right) \quad (4)$$

where $\hat{\lambda}_{N,j}$ is set to 0 for $j > N$. Convergence to zero of the right-hand side is the condition under which the algorithm provides a uniform approximation to the cdf of $\sum_{j=1}^{\infty} \lambda_j Z_j^2$.

Now, $\sum_{j=1}^N \hat{\lambda}_{N,j} Z_j^2$ is a quadratic form in normal random vectors. Indeed, setting $\mathbf{Z} := (Z_1, \dots, Z_N)'$ and $\hat{\Lambda}_N := \text{dg}(\hat{\lambda}_{N,1}, \dots, \hat{\lambda}_{N,N})$ ($\text{dg}(\mathbf{v})$ is the matrix having \mathbf{v} on its diagonal and 0's elsewhere), it follows that $\sum_{j=1}^N \hat{\lambda}_{N,j} Z_j^2 = \mathbf{Z}' \hat{\Lambda}_N \mathbf{Z}$. The quantiles of the distribution can be obtained using a root finding algorithm, as solutions of the equation $\mathbb{P} \left(\sum_{j=1}^N \hat{\lambda}_{N,j} Z_j^2 \leq x \right) = p$. In particular, even a simple algorithm such as the bisection method appears in this case to be a good choice. This is discussed in Appendix A.1.

In the following, it will be shown how the first step, that is the replacement of the integral operator through an approximate one, can be performed with the so-called Wielandt–Nyström method, by using different quadrature rules that can be more or less suited to the computation of eigenfunctions and eigenvalues.

3.1. The Wielandt–Nyström method

In this section, the Wielandt–Nyström method for a kernel $h(\cdot, \cdot)$ defined on $[0, 1] \times [0, 1]$ is presented. The method extends with minor modifications to the case in which $[0, 1]$ is replaced by a more general set. Moreover, the probability \mathbb{P} is supposed to be the uniform distribution, a fact that can generally be obtained through the probability integral transform.

Consider the operator \mathcal{H} defined as:

$$\mathcal{H}\phi(x) = \int_0^1 h(x, y) \phi(y) dy, \quad (5)$$

for $x \in [0, 1]$, $\phi \in L^2$. Let $(\lambda_j)_j$ and $(\phi_j)_j$ be respectively the eigenvalues and eigenfunctions of the operator, whose existence is guaranteed under the conditions leading to (2).

The Wielandt–Nyström method replaces the integral with an approximate one using a quadrature rule given by the sequence of nodes $(y_{N,k})_{k=1, \dots, N}$ and weights $(w_{N,k})_{k=1, \dots, N}$. To ease the notation, the subscript N will be omitted and $(y_k)_{k=1, \dots, N}$ and $(w_k)_{k=1, \dots, N}$ will be used instead of $(y_{N,k})_{k=1, \dots, N}$ and $(w_{N,k})_{k=1, \dots, N}$. All the quadrature rules considered are such that $w_k \geq 0$ for $k = 1, \dots, N$ and $\sum_{k=1}^N w_k = 1$, the length of the interval $[0, 1]$.

Then, the integral operator \mathcal{H} of (5) is replaced by the approximate one given by:

$$\mathcal{H}_N \phi(x) = \sum_{k=1}^N h(x, y_k) \phi(y_k) w_k.$$

One can solve the equation $\lambda \phi(x) = \mathcal{H}_N \phi(x)$ through the linear system:

$$\lambda \phi(y_i) = \mathcal{H}_N \phi(y_i) = \sum_{k=1}^N h(y_i, y_k) \phi(y_k) w_k,$$

where $y_i \in (y_k)_{k=1, \dots, N}$. Thus one can get N eigenvalues by solving the matricial eigenvalue problem $\det(\lambda \mathbf{I}_N - \tilde{\mathbf{H}}_N)$ where $\tilde{\mathbf{H}}_N$ has generic element $\tilde{h}_{ik,N} = h(y_i, y_k) w_k$ and \mathbf{I}_N is the identity matrix of size N . To exploit the greater precision of eigenvalue computations in the symmetric case, a new matrix $\mathbf{D}_N = \text{dg}(w_1, \dots, w_N)$ is introduced and, instead of the non-symmetric $\tilde{\mathbf{H}}_N$, one considers the alternative eigenvalue problem:

$$\det \left(\lambda \mathbf{I}_N - \mathbf{D}_N^{\frac{1}{2}} \tilde{\mathbf{H}}_N \mathbf{D}_N^{-\frac{1}{2}} \right) = \det(\lambda \mathbf{I}_N - \mathbf{H}_N)$$

where $\mathbf{A}^{\frac{1}{2}}$ is a square root of \mathbf{A} and $\mathbf{A}^{-\frac{1}{2}}$ is its inverse. The spectra of $\tilde{\mathbf{H}}_N$ and \mathbf{H}_N coincide but the replacement improves the performance of the numerical procedure as now \mathbf{H}_N is symmetric with generic element given by $h_{ik,N} = h(y_i, y_k) \sqrt{w_i w_k}$.

The possibility of setting the diagonal of the matrix \mathbf{H}_N to zero, as in Koltchinskii and Giné (2000, Eq. (1.3)), has also been considered. In a manner similar to what happens with the asymptotic theory of quadratic statistics, this choice provides results under less restrictive assumptions (see Koltchinskii and Giné, 2000, p. 123, after Corollary 3.3) but this seems to worsen the error in the computation of the eigenvalues and eigenfunctions. In any case, in Appendix A.3 it is shown that the L^2 -distance of the spectra between the case when the diagonal is set of zero and when it is not is usually $O(N^{-\frac{1}{2}})$ (or $O_{\mathbb{P}}(N^{-\frac{1}{2}})$ for the Monte Carlo method).

The set of eigenvalues $(\hat{\lambda}_{N,j})_{j=1,\dots,N}$ of the matrix \mathbf{H}_N approximates the first N eigenvalues $(\lambda_j)_{j=1,\dots,N}$ of \mathcal{H} . If $\hat{\phi}_{N,j}$ is the j -th right eigenvector of \mathbf{H}_N , the j -th right eigenvector of $\tilde{\mathbf{H}}_N$ is given by $\tilde{\phi}_{N,j} = \mathbf{D}_N^{-\frac{1}{2}} \hat{\phi}_{N,j}$. Indeed, it is sufficient to premultiply $(\hat{\lambda}_{N,j} \mathbf{I}_N - \tilde{\mathbf{H}}_N) \tilde{\phi}_{N,j} = \mathbf{0}$ by $\mathbf{D}_N^{\frac{1}{2}}$ and use $\mathbf{H}_N = \mathbf{D}_N^{\frac{1}{2}} \tilde{\mathbf{H}}_N \mathbf{D}_N^{-\frac{1}{2}}$ to get $(\hat{\lambda}_{N,j} \mathbf{I}_N - \mathbf{H}_N) \mathbf{D}_N^{\frac{1}{2}} \tilde{\phi}_{N,j} = 0$, from which $\hat{\phi}_{N,j} = \mathbf{D}_N^{\frac{1}{2}} \tilde{\phi}_{N,j}$ and $\tilde{\phi}_{N,j} = \mathbf{D}_N^{-\frac{1}{2}} \hat{\phi}_{N,j}$.

If $[\mathbf{x}]_k$ denotes the k -th element of the vector \mathbf{x} , the eigenfunctions can then be computed using the so-called *Nyström extension* or *interpolation* formula:

$$\hat{\phi}_{N,j}(x) = \frac{1}{\hat{\lambda}_{N,j}} \sum_{k=1}^N h(x, y_k) w_k [\tilde{\phi}_{N,j}]_k. \quad (6)$$

As $\hat{\phi}_{N,j}(y_k) = [\tilde{\phi}_{N,j}]_k$, the formula provides a function passing through the points $(y_k, [\tilde{\phi}_{N,j}]_k)_{k=1,\dots,N}$, hence the name interpolation, but it extends the function to other values of x , hence the name extension.

A reminder of the relevant formulas is provided in Algorithm 1. In the following, some hints about the computational complexity of each step in terms of operations, i.e. in the real-number model of computation, are provided. The real-number model describes a situation in which algebraic operations, namely addition, subtraction, multiplication, division, raising to fractional powers, as well as comparisons between numbers, are performed with infinite precision at unit cost. This is called algebraic complexity in Borwein and Borwein (1988, p. 591). The complexity of the different steps can be described as follows:

- The computational complexity of the determination of nodes and weights depends on the choice of the quadrature rule. As an example, the N -point Clenshaw–Curtis formula can be implemented in $O(N \ln N)$ operations through the Fast Fourier Transform. The N -point Gauss–Legendre formula requires $O(N^2)$ operations by solving a tridiagonal eigenvalue problem, or $O(N^3)$ when the tridiagonal structure of the eigenvalue problem is neglected (see Trefethen, 2008).
- If the kernel h contains only algebraic operations, populating the matrix \mathbf{H}_N takes $O(N^2)$ operations. If the kernel h contains non-algebraic functions, the computational complexity of these functions must be accounted for. As an example, the Anderson–Darling kernel contains a logarithm that can be evaluated with error ε in $O(|\ln \varepsilon|)$ operations, therefore, populating the matrix requires $O(N^2 |\ln \varepsilon|)$ operations.
- Computing the eigendecomposition can be performed as fast as matrix multiplication (see Pan and Chen, 1999 and Demmel et al., 2007), i.e. through $O(N^{\omega+\eta})$ operations for any $\eta > 0$, where the best value of ω is unknown, but is larger than or equal to 2. The fastest known algorithm (see Alman and Vassilevska Williams, 2021) has $\omega < 2.3728596$ but is not used in practice. Other algorithms have $\omega = \log_2 7$ (Strassen algorithm) or $\omega = 3$. Therefore, this step takes at most $O(N^{3+\eta})$ operations for any $\eta > 0$.
- The computation of the Nyström extension formula require $O(N)$ operations per value of x and per eigenfunction. If the kernel h contains non-algebraic functions, this should be accounted for as above.
- At last, the computational complexity of the Davies algorithm is very difficult to characterize.

This implies that, apart from the Davies algorithm, the overall complexity is dictated by the construction of the quadrature rule and by the eigendecomposition, and is $O(N^{\omega+\eta})$, for $2 \leq \omega \leq 3$ and any $\eta > 0$. If raising to fractional powers is excluded from the list of operations with unit cost, the relevant concept of complexity is rational complexity (see Borwein and Borwein, 1988, p. 591). In this case, the computation of $\sqrt{w_k}$ and $\sqrt{w_i w_k}$, as in steps 2 and 5 of Algorithm 1, with error ε requires $O(|\ln \varepsilon|)$ operations.

The choice of the nodes $(y_j)_{j=1,\dots,N}$ and the weights $(w_j)_{j=1,\dots,N}$ depends on the objective of the analysis and on the characteristics of the kernel h . For kernels defined on $[0, 1] \times [0, 1]$ or on $\mathbb{R} \times \mathbb{R}$, Gaussian quadrature rules are the most common choice; see Press et al. (2002, Chapter 18) or Atkinson (1976). The reason is that the convergence rate of Gaussian quadrature is exponential or super-exponential.

However, as shown below, Gaussian quadrature is not always the optimal choice for statistical applications.

A first possible choice is to use Monte Carlo sampling: this is advocated by Koltchinskii and Giné (2000) for the computation of the eigenvalues of an integral operator, but the objective of that paper is intrinsically different from the one of this paper (see also Adamczak and Bednorz, 2015, where iid points are replaced by a realization of a Markov chain). However, as shown in the following, Monte Carlo has quite bad convergence properties.

Algorithm 1: Wielandt–Nyström algorithm and computation of the cdf.

-
- Input:** $h(\cdot, \cdot)$
- Output:** $(\hat{\lambda}_{N,j})_{j=1,\dots,N}, (\hat{\phi}_{N,j}(\cdot))_{j=1,\dots,N}$
- 1 Generate the nodes $(y_k)_{k=1,\dots,N}$ and weights $(w_k)_{k=1,\dots,N}$ of the quadrature rule
 - 2 Populate the matrix \mathbf{H}_N with (i, j) -element $h_{ik,N} = h(y_i, y_k) \sqrt{w_i w_k}$
 - 3 Compute the eigenvalues $(\hat{\lambda}_{N,j})_{j=1,\dots,N}$ of \mathbf{H}_N
 - 4 Compute the normalized eigenvectors $\hat{\phi}_{N,j}$ of \mathbf{H}_N
 - 5 For each j and for $x \in [0, 1]$, compute the eigenfunction $\hat{\phi}_{N,j}(x) = \frac{1}{\hat{\lambda}_{N,j}} \sum_{k=1}^N h(x, y_k) \sqrt{w_k} [\hat{\phi}_{N,j}]_k$
 - 6 Compute the cdf of the random variable $\sum_{j=1}^N \hat{\lambda}_{N,j} Z_j^2$ through the Davies algorithm
-

An alternative is to replace Gaussian quadrature rules with quasi-Monte Carlo sequences (see [Choirat and Seri, 2006](#)). Even if the rate of convergence is often worse than for Gaussian quadrature rules, quasi-Monte Carlo points are simpler to obtain. Therefore, for a low-discrepancy sequence $(y_k)_{k=1,\dots,N}$ with weights equal to $\frac{1}{N}$, the set of eigenvalues $(\hat{\lambda}_{N,j})_{j=1,\dots,N}$ is obtained as the spectrum of the matrix \mathbf{H}_N with generic element $h_{ik,N} = \frac{1}{N} h(y_i, y_k)$. The results in Section 5.1 show that quasi-Monte Carlo works well for the computation of the spectrum associated to quadratic statistics.

Another very simple numerical integration technique is the trapezium rule. This corresponds to the sequence of points $(y_k)_{k=1,\dots,N} = \left(\frac{k-1}{N-1} \right)_{k=1,\dots,N}$ and weights $(w_k)_{k=1,\dots,N} = \left(\frac{1}{2(N-1)}, \frac{1}{N-1}, \dots, \frac{1}{N-1}, \frac{1}{2(N-1)} \right)$.

At last, some recent contributions, and in particular [Trefethen \(2008\)](#) and [Xiang and Bornemann \(2012\)](#), have shown that the Clenshaw–Curtis quadrature rule has often similar convergence properties to Gauss quadrature but has an advantage when it comes to the computation of the nodes and the weights. It may be useful to point out that the so-called Clenshaw–Curtis quadrature rule should be more correctly called Fejér's first rule ([Fejér, 1933](#)). The difference among Clenshaw–Curtis, Fejér's first and Fejér's second rules are described in [Waldvogel \(2006\)](#) and [Trefethen \(2008\)](#), where it is also stated that Fejér's first rule is sometimes called “classical” Clenshaw–Curtis rule. In the following, Fejér's first rule will be used instead of Clenshaw–Curtis rule as it can also be applied to kernels (such as the Anderson–Darling one) that diverge at the endpoints of the interval. However, for coherence with the literature, the name of Clenshaw–Curtis rule is adopted here.

4. Theoretical results

In this section, some theoretical results concerning the convergence and the approximation error of eigenvalues, eigenfunctions and the cdf are discussed.

4.1. Results on eigenvalues

The Nyström method can be applied in two different situations. First, it can be used for the numerical solution of Fredholm integral equations, defined as $\mathcal{H}m(x) = \lambda m(x)$ where \mathcal{H} is an integral operator, $\lambda \neq 0$ is known and the objective is to obtain the function m . Second, it can be used to find the eigenvalues and the eigenfunctions of \mathcal{H} . The second situation is the one considered here.

When applied to the numerical solution of Fredholm integral equations, properties of the Nyström method have often been investigated using the theory of collectively compact operators; see, e.g., [Anselone \(1971\)](#), [Linz \(1979\)](#), [Atkinson and Han \(2005\)](#). As it concerns the properties of the kernels required for the approximation to converge, it is enough to reason as in [Linz \(1979, pp. 177–178\)](#). A first very simple set of conditions is given by boundedness and continuity of the kernel h . In [Anselone \(1967\)](#) and [Atkinson \(1967a\)](#), it is shown respectively how the theory can be generalized to cover kernels that are bounded and discontinuous, and weakly singular.

When dealing with the computation of the eigenvalues and the eigenfunctions of \mathcal{H} , the problem can be seen as a two-step one: first of all, a sequence of approximate eigenvalues is computed and then a set of Fredholm equations is solved, in which the known constants are replaced by the approximate eigenvalues. Convergence results for the approximated eigenvalues of \mathcal{H} have been proved in [Hilbert \(1904, 1912\)](#), where a special integration rule is used, and [Bückner \(1950, p. 110, 1952, p. 362\)](#); note that [Wielandt, 1956](#), p. 263, inverts the pages of the two sources), where the result is extended to more general rules. Here a result from [Wielandt \(1956\)](#), as stated in [Rakotch \(1975, p. 795\)](#), is reported.

In the statements of the results, the following notation will be used. Let $\hat{\lambda}_{N,1}^+ \geq \hat{\lambda}_{N,2}^+ \geq \dots \geq \hat{\lambda}_{N,r}^+ > 0 > \hat{\lambda}_{N,s}^- \geq \dots \geq \hat{\lambda}_{N,2}^- \geq \hat{\lambda}_{N,1}^-$ be the r largest positive and the s smallest negative eigenvalues of \mathcal{H}_N , and let $\lambda_1^+ \geq \lambda_2^+ \geq \dots \geq \lambda_r^+ > 0 > \lambda_s^- \geq \dots \geq \lambda_2^- \geq \lambda_1^-$ be the corresponding eigenvalues of \mathcal{H} . Moreover:

$$\eta_N(x, y) := \sum_{k=1}^N w_k h(x, y_k) h(y_k, y) - \int_0^1 h(x, z) h(z, y) dz. \quad (7)$$

An important property is that the quadrature rule is convergent with respect to $h(x, y)$, i.e. $\eta_N(x, y)$ converges to 0 uniformly in $[0, 1] \times [0, 1]$ or $\|\eta_N\|_\infty \downarrow 0$.

Theorem 1. Let $h(x, y)$ be a Hermitian kernel on $[0, 1] \times [0, 1]$, i.e. $h(x, y) = \overline{h(y, x)}$, such that $x \mapsto \int_0^1 |h(x, y)|^2 dy$ is bounded on $[0, 1]$. Suppose that $\|\eta_N\|_\infty \downarrow 0$ and $\sum_{j=1}^N w_j = 1$. Then:

$$\lambda_i^+ = \lim_{N \rightarrow \infty} \hat{\lambda}_{N,i}^+, \quad \lambda_j^- = \lim_{N \rightarrow \infty} \hat{\lambda}_{N,j}^-, \quad i = 1, \dots, r, j = 1, \dots, s,$$

and this convergence is uniform in i and j , i.e. $|\hat{\lambda}_{N,k} - \lambda_k| \leq q_N$ with $\lim_{N \rightarrow \infty} q_N = 0$, where either $\hat{\lambda}_{N,k} = \hat{\lambda}_{N,k}^+$, $\lambda_k = \lambda_k^+$ or $\hat{\lambda}_{N,k} = \hat{\lambda}_{N,k}^-$, $\lambda_k = \lambda_k^-$.

In the following some hints about the behavior of $|\hat{\lambda}_{N,j} - \lambda_j|$ are provided. One should realize that the results presented in the literature are not necessarily comparable to each other. First, some results provide error estimates, i.e. bounds that are amenable to (approximate) computation and can be used to investigate the precision of the algorithm, while others concern convergence rates, i.e. the order of decrease of $|\hat{\lambda}_{N,j} - \lambda_j|$ with N and j . The former are often cruder as they must contain only computable (or approximable) quantities. Second, some results are stated in terms of the distance between an eigenvalue of \mathcal{H} and the nearest eigenvalue of \mathcal{H}_N , while others involve the distance between the j -th eigenvalues of \mathcal{H} and \mathcal{H}_N , both counted with their multiplicity. The main difference between the two situations arises for multiple eigenvalues: when results are stated for an eigenvalue of \mathcal{H} and the nearest eigenvalue of \mathcal{H}_N , they do not take into account the potential difference in the multiplicity of the eigenvalues.

The most complete results seem to be the ones exploiting the theory of prolongation and restriction operators: it turns out that the approximated eigenvalues converge to the true ones (under hypotheses (I)-(VI) on pp. 58 and 60 in Spence, 1975), both if eigenvalues are simple and multiple (see respectively Theorems 3 and 7 in Spence, 1975); the extension to the case of unbounded kernels is covered in Section 4 of the same paper. A complete study of error estimates for eigenvalues (and eigenfunctions) in the case of kernels with simple eigenvalues is in Spence (1977, see p. 140 for the case of multiple eigenvalues).

However, the present paper aims at giving some hints about convergence rates for eigenvalues rather than providing error estimates. In Keller (1965, Corollary 2.1), pointwise bounds for the distance between an eigenvalue of \mathcal{H} and the nearest eigenvalue of \mathcal{H}_N (evaluated through a quadrature rule with positive weights) are provided. In the case of simple eigenvalues, these bounds can be used to provide a convergence rate since any approximated eigenvalue converges to its real value and, for N large enough, the approximated eigenvalue that is the nearest to λ_j is the j -th one, that is $\hat{\lambda}_{N,j}$; as explained above, the only situation that raises some problem arises when eigenvalues are multiple. Therefore, for simple eigenvalues:

$$|\hat{\lambda}_{N,j} - \lambda_j| \lesssim \sup_{y_m \in (y_k)_{k=1, \dots, N}} \left| \sum_{k=1}^N w_k h(y_m, y_k) \phi_j(y_k) - \int_0^1 h(y_m, z) \phi_j(z) dz \right| \quad (8)$$

(note that the original source contains a typo, as u_k in the right-hand side of equation (33) should be u_j , and defines eigenvalues in a non-standard way). The behavior in N of this term is dictated by the integration error (as an example, for Monte Carlo integration it behaves like $O_{\mathbb{P}}(N^{-\frac{1}{2}})$, etc.); however, the approximation error also depends on j through the fact that higher-order eigenfunctions oscillate more rapidly or equivalently that their derivatives grow in absolute value with j .

The following theorem reproduces the bound derived in Rakotch (1975) involving the j -th eigenvalue of \mathcal{H} and the j -th eigenvalue of \mathcal{H}_N for $1 \leq j \leq N$, for simple and multiple eigenvalues. Moreover, a corollary that simplifies the bound is stated after the theorem.

Theorem 2. Let $h(x, y)$ be a Hermitian kernel on $[0, 1] \times [0, 1]$, i.e. $h(x, y) = \overline{h(y, x)}$, such that $x \mapsto \int_0^1 |h(x, y)|^2 dy$ is bounded on $[0, 1]$. Let:

$$\gamma_N := \left[\sum_{j=1}^N w_j \int_0^1 |\eta_N(x, y_j)|^2 dx \right]^{\frac{1}{2}},$$

$$\alpha_N := \left[\int_0^1 \int_0^1 |\eta_N(x, y)|^2 dxdy \right]^{\frac{1}{2}},$$

$$\beta_N := \left[\sum_{i=1}^N \sum_{j=1}^N w_i w_j |\eta_N(y_i, y_j)|^2 \right]^{\frac{1}{2}}$$

and:

$$\rho_N := \max \{\alpha_N, \beta_N\}.$$

Then, if $\|\eta_N\|_\infty \downarrow 0$ and $\sum_{j=1}^N w_j = 1$, the following two bounds hold:

- For any j :

$$|\hat{\lambda}_{N,j} - \lambda_j| \leq \left(\frac{1+\sqrt{5}}{2} (\gamma_N + \rho_N) \right)^{\frac{1}{2}}.$$

- If $\lambda_j^2 \vee \hat{\lambda}_{N,j}^2 \geq C\rho_N$ for some $C > 1$, then:

$$|\hat{\lambda}_{N,1} - \lambda_1| \leq \frac{\gamma_N}{(\lambda_1^2 \vee \hat{\lambda}_{N,1}^2 - \rho_N)^{\frac{1}{2}}},$$

$$|\hat{\lambda}_{N,j} - \lambda_j| \leq \frac{\gamma_N + \rho_N}{(\lambda_j^2 \vee \hat{\lambda}_{N,j}^2 - \rho_N)^{\frac{1}{2}}}.$$

Remark. For $j > 1$, the second bound is better than the first one when:

$$\lambda_j^2 \vee \hat{\lambda}_{N,j}^2 \geq \frac{2}{1+\sqrt{5}} \gamma_N + \frac{1+\sqrt{5}}{2} \rho_N.$$

This implies that taking the minimum of the bounds automatically satisfies the requirement that $\lambda_j^2 \vee \hat{\lambda}_{N,j}^2 \geq C\rho_N$ for some $C > 1$.

Corollary 1. Under the conditions of Theorem 2:

$$|\hat{\lambda}_{N,j} - \lambda_j| \leq \begin{cases} \|\eta_N\|_\infty (\lambda_1^2 - \|\eta_N\|_\infty)^{-\frac{1}{2}}, & j = 1, \lambda_1^2 \geq \frac{3+\sqrt{5}}{4} \|\eta_N\|_\infty, \\ (1+\sqrt{5})^{\frac{1}{2}} \|\eta_N\|_\infty^{\frac{1}{2}}, & j = 1, \lambda_1^2 < \frac{3+\sqrt{5}}{4} \|\eta_N\|_\infty, \\ 2 \|\eta_N\|_\infty (\lambda_j^2 - \|\eta_N\|_\infty)^{-\frac{1}{2}}, & j > 1, \lambda_j^2 \geq \sqrt{5} \|\eta_N\|_\infty, \\ (1+\sqrt{5})^{\frac{1}{2}} \|\eta_N\|_\infty^{\frac{1}{2}}, & j > 1, \lambda_j^2 < \sqrt{5} \|\eta_N\|_\infty. \end{cases}$$

Even if the bounds of Theorem 2 and Corollary 1 appear more complicated than the one in (8), it is simple to see that for bounded and continuous kernels also this bound is decreasing in N with a rate given by the integration error but is increasing in j because of the effect of λ_j and $\hat{\lambda}_{N,j} \rightarrow \lambda_j$. Roughly speaking, $|\hat{\lambda}_{N,j} - \lambda_j| = O(\|\eta_N\|_\infty / |\lambda_j|)$, so that higher-order eigenvalues are approximated with larger error, a fact that is observed in practice. Note that this bound eliminates the dependence on the unknown eigenfunction ϕ_j .

4.2. Results on eigenfunctions

As concerns the eigenfunctions, the results are less complete and more complex than for eigenvalues. On the one hand, eigenfunctions are more difficult to study with respect to eigenvalues. As they are elements of a functional space, results differ according to the kind of convergence (pointwise, uniform, etc.) that is used. On the other hand, a part of the error in their approximation comes from the previous approximation of the eigenvalues and the results about that, as shown above, are far from easy. This problem has been studied in Atkinson (1967b,a) (using the theory of collectively compact operators), Thomas (1974), Spence (1975, 1977, 1979) (using the theory of prolongation and restriction operators), Keller (1965) and Rakotch (1975). However, the discussion in Spence (1977) points at a general equivalence between the rate of convergence of eigenvalues and of eigenfunctions.

When the objective is the computation of an integral involving the eigenfunctions, it is often better to use the approximated eigenvector instead of the approximated eigenfunction. As an example, the coefficients $a_j = \int g\phi_j d\mathbb{P}$ for $j = 1, \dots$, appearing in the asymptotic distribution of V_n under a Pitman drift, can be approximated as:

$$\hat{a}_{N,j} = \sum_{k=1}^N w_k g(y_k) [\tilde{\phi}_{N,j}]_k.$$

It would be possible to compute these quantities using the function ϕ_j , as approximated through formula (6), evaluated for a set of points different from $(y_k)_{k=1,\dots,N}$. The rationale is that, while the number of points used for computing the eigenvalues and the eigenvectors is bounded by computational requirements, the number of points used in the integration is less critical from the point of view of resource scarcity. However, there is a practical reason for performing the integration only at the points $(y_k)_{k=1,\dots,N}$. Indeed, many methods for the solution of integral equations are *superconvergent* (see, e.g., Atkinson and Han, 2005, p. 475), i.e. they have a convergence rate at the nodes $(y_k)_{k=1,\dots,N}$ that is faster than the convergence rate over the rest of the interval. This implies that using the function at points not belonging to $(y_k)_{k=1,\dots,N}$ may need a much larger amount of points.

4.3. Results on the cdf

Whenever the objective of the analysis is the computation of the cdf, it is useful to choose N in such a way to keep the approximation error under a certain threshold. Unfortunately, the derivation of error bounds and convergence rates for the computation of the cdf is difficult. Indeed, while bounds for the uniform distance between the distributions of two weighted sums of χ^2 random variables exist (see, e.g., (4)), the bounds on the eigenvalues in Theorem 2 and Corollary 1 are hardly precise enough to be used to derive convergence rates. Some bounds for the overall approximation of the spectrum, i.e. for $\sqrt{\sum_{j=1}^{\infty} (\hat{\lambda}_{N,j} - \lambda_j)^2}$ (see Section 5.2), may be obtained from Section 4 of Koltchinskii and Giné (2000) but, even in the simplest case, the bounds they provide are far from optimal (see Appendix A.3). Moreover, bounds obtained replacing the results of Theorem 2 and Corollary 1 into (4) can be misleading. As an example, according to the bounds in Theorem 2 and Corollary 1, the error $|\hat{\lambda}_{N,j} - \lambda_j|$ with large subscript j can be so important that it may seem better to set the approximate eigenvalue to 0, rather than using $\hat{\lambda}_{N,j}$. This suggests to approximate $\sum_{j=1}^{\infty} \lambda_j Z_j^2$ through $\sum_{j=1}^{N'} \hat{\lambda}_{N,j} Z_j^2$ with $N' < N$. However, some anecdotal evidence (see Choirat and Seri, 2013; Seri, 2017) shows that the approximation of $\sum_{j=1}^{\infty} \lambda_j Z_j^2$ is much better when $\sum_{j=1}^{N'} \hat{\lambda}_{N,j} Z_j^2$ has the same expectation of $\sum_{j=1}^{\infty} \lambda_j Z_j^2$. Therefore, especially when $\sum_{j=1}^{\infty} \lambda_j = \sum_{j=1}^N \hat{\lambda}_{N,j}$, neglecting the smallest approximated eigenvalues can lead to an increase of the error. This does not rule out the possibility that neglecting the smallest approximated eigenvalues improves the approximation of the distribution of $\sum_{j=1}^{\infty} \lambda_j (Z_j^2 - 1)$, because in this case centering is automatic. As the reader may see, the situation is quite complex. For this reason, this topic is left for a future paper. For the time being, the computations show that a larger N is always better and there is no value added from taking $N' < N$.

5. Computational results

The algorithm previously described has been implemented in an R script. The computation of the cdf uses an algorithm implemented in a package described in Duchesne and Lafaye De Micheaux (2010).

The following approximate integration methods will be used:

MC	Monte-Carlo method. N is equal to the number of pseudo-random points drawn in $[0, 1]$.
HA	Integration through the one-dimensional Halton (or van der Corput) sequence.
TR	Trapezium rule.
GL	Gauss-Legendre quadrature rule.
CC	Clenshaw-Curtis quadrature rule.

These quadrature rules have been applied to three kernels described in Table 1. The choice of these kernels was dictated by the fact that they represent a sufficiently broad scope of possible applications: the Cramér-von Mises kernel has simple eigenvalues and is bounded, the Anderson-Darling kernel has simple eigenvalues and is unbounded, the Watson kernel has multiple eigenvalues. In the theory outlined in Section 4, bounded kernels with simple eigenvalues are considered as the baseline case, while deviations from this case have to be taken into account on a case by case basis.

Some comments on the quadrature rules may be necessary. First, HA corresponds to what is called a sequence in the relevant literature (see Niederreiter, 1992, p. 14): the name “sequence” is used to denote a finite segment extracted from an infinite sequence, so that $(y_k)_{k=1,\dots,N}$ is obtained from $(y_k)_{k=1,\dots,N-1}$ by adding a single point, while the name “point set” is reserved for the cases in which a different set of points has to be computed for any value of N . It is clear that sequences are easier to compute but often less performing than point sets. In this case, the one-dimensional van der Corput sequence, a low-discrepancy sequence whose n -th element is constructed by a symmetric reflection around the decimal point of the digit expansion of the natural number n in base b (see Niederreiter, 1992, p. 25), has been used. The sequence depends on the choice of a base b : in the applications, $b = 3$ has been taken because, despite the uniformity properties of van der Corput sequences are known to deteriorate when b increases, the minimum value of the star discrepancy is obtained for

Table 1

The kernels used in the computations (here P_j denotes the j -th Legendre polynomial and $\lfloor \cdot \rfloor$ denotes the floor function).

Name	Kernel	Eigenvalue λ_j	Eigenfunction $\phi_j(x)$
Cramér–von Mises	$\frac{x^2}{2} + \frac{y^2}{2} - x \vee y + \frac{1}{3}$	$j^{-2}\pi^{-2}$	$\sqrt{2}\cos(\pi jx)$
Anderson–Darling	$-\ln(x \vee y - xy) - 1$	$j^{-1}(j+1)^{-1}$	$\sqrt{2j+1}P_j(2x-1)$
Watson	$\frac{(x-y)^2}{2} + \frac{x+y}{2} - x \vee y + \frac{1}{12}$	$\frac{1}{4\pi^2} \left\lfloor \frac{j+1}{2} \right\rfloor^{-2}$	$\begin{cases} \sqrt{2}\sin(\pi(j+1)x), & j \text{ odd}, \\ \sqrt{2}\cos(\pi jx), & j \text{ even} \end{cases}$

$b = 3$ (see Niederreiter, 1992, p. 25). The results are quantitatively, but not qualitatively, different for other values of b . Moreover, this sequence is referred to as a Halton sequence in the following, because this is the name that is used in the multidimensional extensions of the van der Corput sequence. Second, the Hammersley point set, a quasi-Monte Carlo set of points that, for $d = 1$, is given by $(y_k)_{k=1,\dots,N} = \left(\frac{2k-1}{2N}\right)_{k=1,\dots,N}$, has also been used; however, as expected, the results were very similar to the ones of the trapezium rule TR, when available. Third, TR cannot be applied to the Anderson–Darling kernel since this quadrature rule requires the kernel to be finite at the extremes of the interval $[0, 1]$.

5.1. Single eigenvalues

The performance of the bound of Corollary 1 is shown here for the Cramér–von Mises statistic. In this case, (7) contains the following quantity:

$$\begin{aligned} & \int_0^1 h(x, z) h(z, y) dz \\ &= -\frac{1}{24} (x-y)^4 - \frac{1}{12} (x+y)^3 - \frac{1}{24} (x-y)^2 + \frac{1}{2} xy (x \wedge y) + \frac{(x \vee y)^3}{6} + \frac{1}{720}. \end{aligned}$$

Figs. 1, 2, 3 and 4 display, respectively for HA, TR, GL and CC, the error in the computation of the eigenvalues, namely $|\hat{\lambda}_{N,j} - \lambda_j|$ for $j = 1, \dots, 9$, and the corresponding bounds of Theorem 2 and Corollary 1. It is apparent that the bounds, especially those of the theorem, are very good for $j = 1$, but they deteriorate rapidly for increasing j . Moreover, for large j and small N the bound does not decrease as $\|\eta_N\|_\infty$, but as $\|\eta_N\|_\infty^{\frac{1}{2}}$, and this does not seem tight.

It should be noted that, if these bounds have to be used to evaluate the performance of the algorithm in the approximation of a specific eigenvalue, the bound of Corollary 1 still holds with $\hat{\lambda}_{N,j}$ replacing λ_j in the right-hand side.

5.2. Spectrum

While the previous bounds are quite effective in bounding the error in the computation of a single eigenvalue, the overall approximation of the spectrum is more difficult to characterize.

The following distance between the spectra of the integral operator and of its finite-dimensional approximation is used:

$$\Lambda^{(2)}(N) = \sqrt{\sum_{j=1}^{\infty} (\hat{\lambda}_{N,j} - \lambda_j)^2}$$

where $\hat{\lambda}_{N,j}$ is set to 0 for $j > N$. This distance is linked to the one considered in Koltchinskii and Giné (2000) and Adamczak and Bednorz (2015). The behavior of this quantity for N varying depends both on the rate of convergence of $\hat{\lambda}_{N,j}$ to λ_j , for $j \leq N$, as a function of N and j , and on the rate of decrease of λ_j to 0, for $j > N$. This distance has been computed for all values of N in the range [10, 1000]. Then, Figs. 5, 6 and 7 display the function $N \mapsto \Lambda^{(2)}(N)$ on a log-log scale for the quadrature methods introduced above. As concerns MC, the figures display both the median (in solid gray line) and the 2.5% and 97.5% quantiles (in dashed gray line) on 1000 replications.

For all the methods, the best possible convergence rate is of the same order as:

$$\Lambda^{(2)}(N) = \sqrt{\sum_{j=1}^N (\hat{\lambda}_{N,j} - \lambda_j)^2 + \sum_{j=N+1}^{\infty} \lambda_j^2} \geq \sqrt{\sum_{j=N+1}^{\infty} \lambda_j^2}.$$

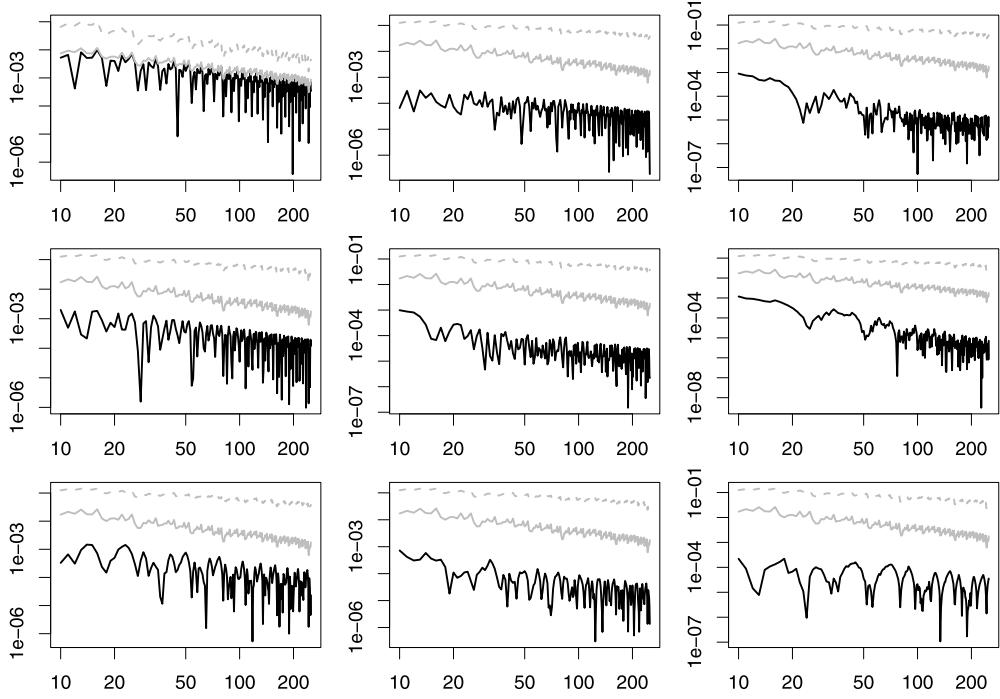


Fig. 1. Distance between the true eigenvalues of the Cramér-von Mises kernel and the ones approximated through HA for varying N : absolute distance $|\hat{\lambda}_{N,j} - \lambda_j|$ (black line), bound from Theorem 2 (gray, solid line) and bound from Corollary 1 (gray, dashed line) for the eigenvalues $j = 1, 2, 3$ (first column, from top to down), $j = 4, 5, 6$ (second column, from top to down), $j = 7, 8, 9$ (third column, from top to down).

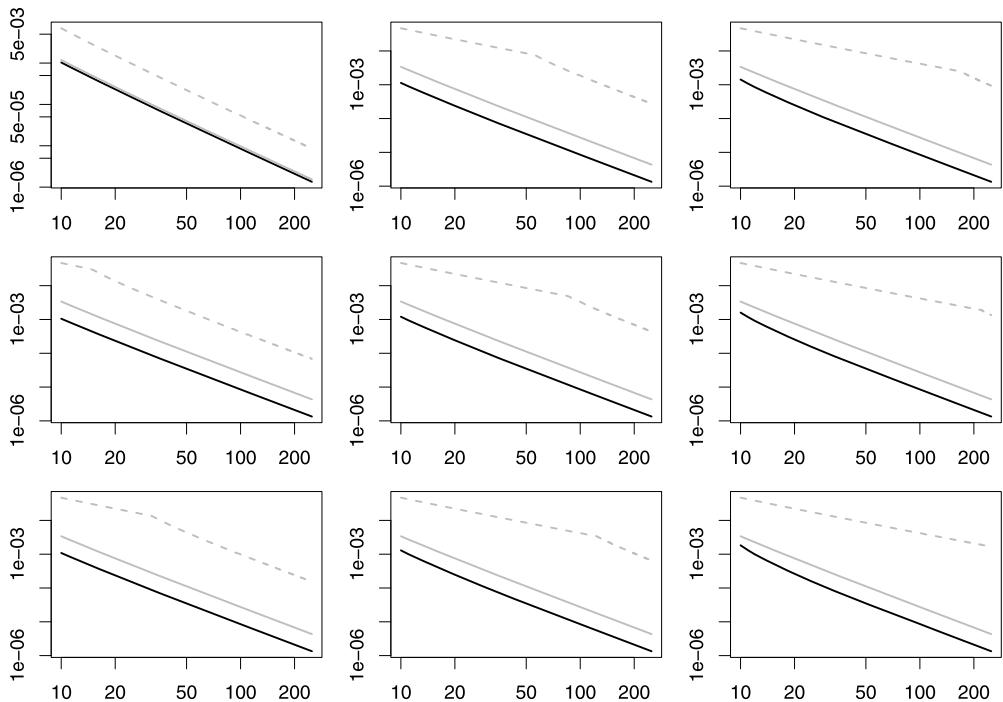


Fig. 2. Distance between the true eigenvalues of the Cramér-von Mises kernel and the ones approximated through TR for varying N : absolute distance $|\hat{\lambda}_{N,j} - \lambda_j|$ (black line), bound from Theorem 2 (gray, solid line) and bound from Corollary 1 (gray, dashed line) for the eigenvalues $j = 1, 2, 3$ (first column, from top to down), $j = 4, 5, 6$ (second column, from top to down), $j = 7, 8, 9$ (third column, from top to down).

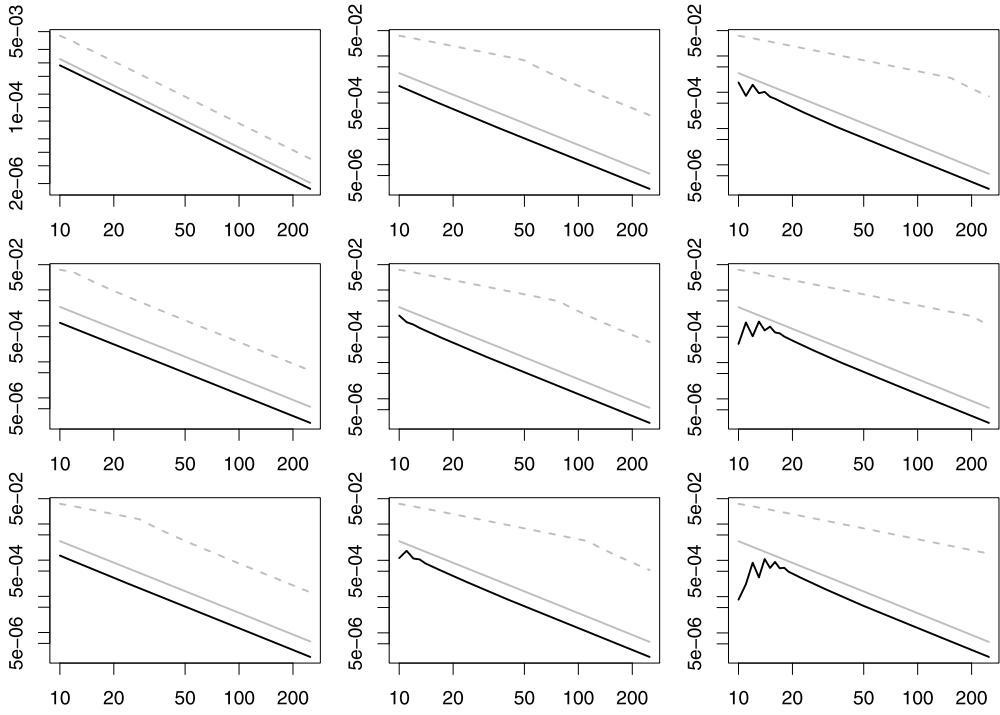


Fig. 3. Distance between the true eigenvalues of the Cramér-von Mises kernel and the ones approximated through GL for varying N : absolute distance $|\hat{\lambda}_{N,j} - \lambda_j|$ (black line), bound from Theorem 2 (gray, solid line) and bound from Corollary 1 (gray, dashed line) for the eigenvalues $j = 1, 2, 3$ (first column, from top to down), $j = 4, 5, 6$ (second column, from top to down), $j = 7, 8, 9$ (third column, from top to down).

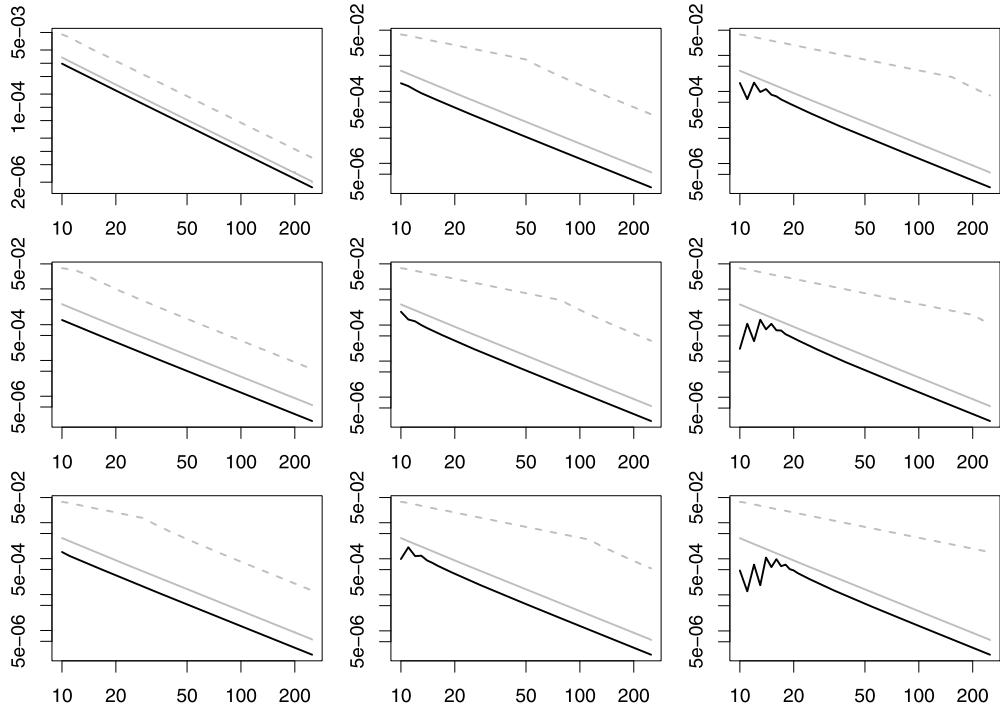


Fig. 4. Distance between the true eigenvalues of the Cramér-von Mises kernel and the ones approximated through CC for varying N : absolute distance $|\hat{\lambda}_{N,j} - \lambda_j|$ (black line), bound from Theorem 2 (gray, solid line) and bound from Corollary 1 (gray, dashed line) for the eigenvalues $j = 1, 2, 3$ (first column, from top to down), $j = 4, 5, 6$ (second column, from top to down), $j = 7, 8, 9$ (third column, from top to down).

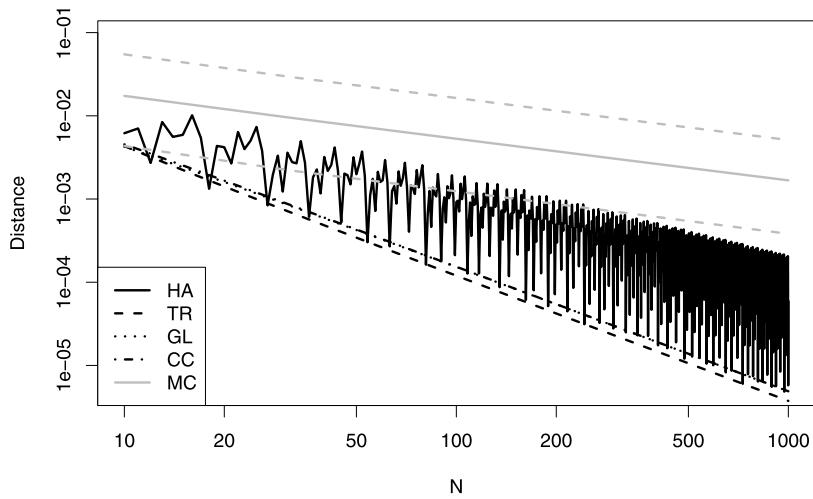


Fig. 5. Distance between the true and the approximated spectrum of the Cramér–von Mises kernel.

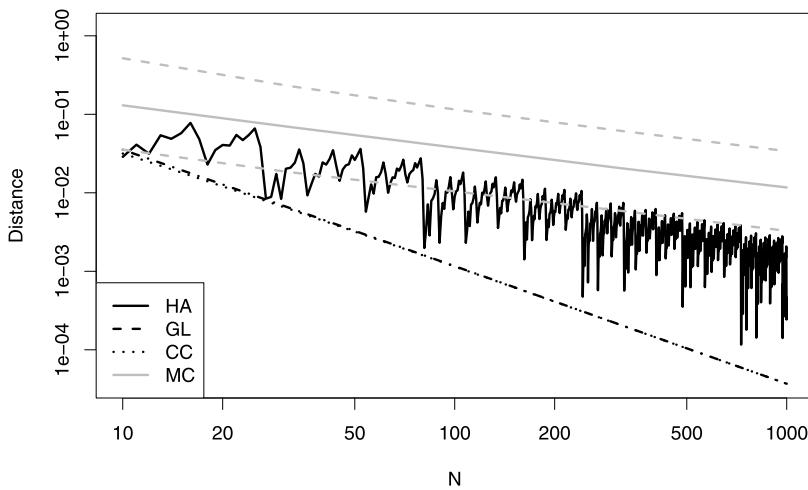


Fig. 6. Distance between the true and the approximated spectrum of the Anderson–Darling kernel.

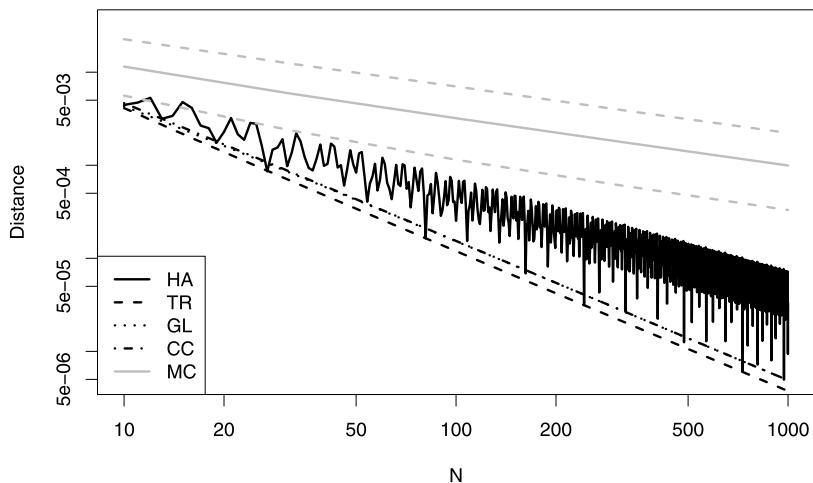


Fig. 7. Distance between the true and the approximated spectrum of the Watson kernel.

If $\lambda_j \sim kj^{-1-\alpha}$ for $\alpha > 0$, $\sum_{j=N+1}^{\infty} \lambda_j^2 \sim k^2 \zeta(2 + 2\alpha, N + 1) \sim \frac{k^2}{2\alpha+1} N^{-2\alpha-1}$ (see Seri, 2015, p. 827), where ζ is the Hurwitz zeta function. Therefore, $\Lambda^{(2)}(N) \gtrsim \frac{k}{\sqrt{2\alpha+1}} N^{-\alpha-\frac{1}{2}}$. For the statistics considered here, $\alpha = 1$ and the best convergence rate is $N^{-\frac{3}{2}}$, that is achieved by TR (when available), GL and CC.

For MC the results of Koltchinskii and Giné (2000, Section 4) can be applied. It turns out that, according to the computations in Appendix A.3, $\Lambda^{(2)}(N) = O_{\mathbb{P}}(N^{-\frac{3}{8}})$ for the Cramér-von Mises and Watson statistics, and $\Lambda^{(2)}(N) = O_{\mathbb{P}}(N^{-\frac{3}{8}} \ln^{\frac{3}{16}} N)$ for the Anderson-Darling statistic. Note, however, that the rate of decrease of $\Lambda^{(2)}(N)$ that is apparent from computational results behaves like $O_{\mathbb{P}}(N^{-\frac{1}{2}})$. This shows that it is difficult to obtain explicit and exact rates of decrease of $\Lambda^{(2)}(N)$ even for the MC case, in which the machinery of probability theory is available.

For HA, let b be the base of the sequence. The segment of the sequence composed of the first $b^s - 1$ elements, for $s \in \mathbb{N}$, corresponds to a rearrangement of the set $(\frac{j}{b^s})_{j=1, \dots, b^s-1}$. The results for HA coincide, for $N = b^s - 1$, with those for the equally spaced grid $(\frac{j}{N+1})_{j=1, \dots, N}$. In practice, this is expected to behave like TR. The performance of HA first deteriorates between the successive values of $(b^s - 1)_{s=1, \dots}$, because the points that are added to the segment of $b^s - 1$ elements worsen the uniformity properties of the sequence up to N , and then improves when b^s approaches $N + 1$.

These results and the computations confirm that MC is always the worst method. HA improves over MC and its performance is better than expected. The general equivalence observed between Gauss-Legendre and Clenshaw-Curtis quadrature rules should not be a surprise according to Trefethen (2008) and Xiang and Bornemann (2012): GL and CC dominate HA and have always the same asymptotic behavior. Moreover, in two cases out of three (namely for the Cramér-von Mises and Watson statistics, that is whenever TR is defined), the rate of convergence of $\Lambda^{(2)}$ to 0 is the same for TR, GL and CC, even if the error is uniformly smaller for TR.

5.3. Eigenfunctions

The Supplementary Material contains plots showing the behavior of the approximated eigenfunctions for the three kernels and the five integration methods. A visual inspection of the graphs shows that the Nyström method, when applied to some of the kernels, gives rise to an unexpected and surprising behavior. Indeed, the approximated eigenfunctions of the Cramér-von Mises and Watson kernels are given by piecewise linear functions, whose knots are the nodes of the quadrature rule. It will be shown below that this happens because the kernels can be represented, on the area $0 \leq x \leq y \leq 1$, as quadratic functions of x and y , and the discontinuity along the diagonal takes a particular form. This does not happen for the Anderson-Darling kernel. Note that this strange fact does not affect the approximation properties of the Nyström extension.

Proposition 1. For any value of $\ell = 1, \dots, N - 1$, the Nyström extension is, for the Cramér-von Mises kernel:

$$\hat{\phi}_{N,j}(x) = \hat{\phi}_{N,j}(y_\ell) - \frac{x - y_\ell}{\hat{\lambda}_{N,j}} \sum_{k=1}^{\ell} w_k [\tilde{\phi}_{N,j}]_k, \quad x \in [y_\ell, y_{\ell+1}),$$

for the Watson kernel:

$$\hat{\phi}_{N,j}(x) = \hat{\phi}_{N,j}(y_\ell) - \frac{x - y_\ell}{\hat{\lambda}_{N,j}} \left\{ \sum_{k=1}^N y_k w_k [\tilde{\phi}_{N,j}]_k + \sum_{k=1}^{\ell} w_k [\tilde{\phi}_{N,j}]_k \right\}, \quad x \in [y_\ell, y_{\ell+1}),$$

and for the Anderson-Darling kernel:

$$\hat{\phi}_{N,j}(x) = \hat{\phi}_{N,j}(y_\ell) + \ln\left(\frac{y_\ell(1-x)}{(1-y_\ell)x}\right) \frac{1}{\hat{\lambda}_{N,j}} \sum_{k=1}^{\ell} w_k [\tilde{\phi}_{N,j}]_k, \quad x \in [y_\ell, y_{\ell+1}).$$

To evaluate the rate of convergence of the eigenfunctions, the L^2 -distance between the first 6 eigenfunctions of the integral operator and their approximations has been computed:

$$\Phi^{(2)}(N) = \sqrt{\sum_{j=1}^6 \int_0^1 (\hat{\phi}_{N,j}(x) - \phi_j(x))^2 dx}.$$

Unfortunately, the correspondence of the approximated eigenfunctions to the real ones cannot be simply obtained using the coupling imposed by the order of appearance of the eigenvalues in the true and the approximated spectrum. First of all,

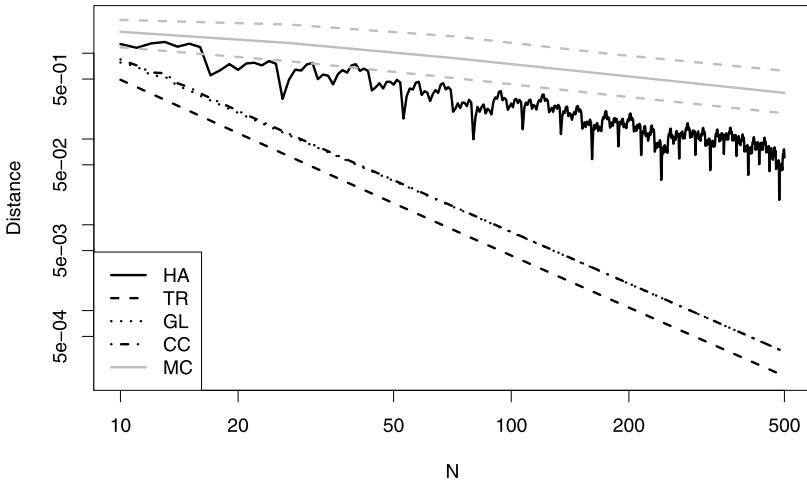


Fig. 8. Distance between the true and the approximated first 6 eigenfunctions of the Cramér–von Mises kernel.

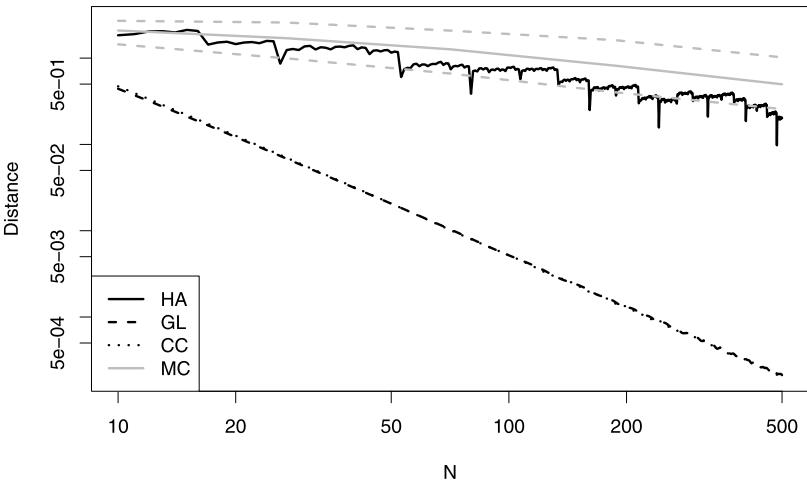


Fig. 9. Distance between the true and the approximated first 6 eigenfunctions of the Anderson–Darling kernel.

eigenfunctions are defined up to their sign. Moreover, eigenvalues often appear in the spectrum with multiple cardinality (this is the case of the Watson kernel, indeed); this implies that more than one approximated and real eigenfunction corresponds to the same eigenvalue. At last, the error of approximation is often larger for the last eigenvalues and this can imply a rank reversal of the eigenvalues and of the associated eigenfunctions. In order to associate an approximated eigenfunction to the corresponding real one, the following algorithm has been used. First of all, the value of j minimizing the quantity $\int_0^1 (\hat{\phi}_{N,j}(x) - \phi_1(x))^2 dx$ has been associated with the first eigenfunction, then, among the remaining values of j , the value of j minimizing $\int_0^1 (\hat{\phi}_{N,j}(x) - \phi_2(x))^2 dx$ has been associated with the second eigenfunction, and so on. The same method has been used also for the graphs in the Supplementary Material. This is not exactly as an overall combinatorial optimization over all the indexes at the same time, but seems to work very well in practice. Also the sign of the approximated eigenfunctions has been chosen in order to minimize the above quantities. The sum has been limited to the first 6 eigenfunctions, since they are sufficiently representative of the behavior of the method and they are not affected by the fluctuations appearing in higher-order eigenfunctions. Moreover, even if this should not be the case, approximated higher-order eigenfunctions can have an imaginary part. However, increasing the number of eigenfunctions does not change the general picture.

The function $N \mapsto \Phi^{(2)}(N)$ for $N \in [10, 500]$ is displayed on a log-log scale in Figs. 8, 9 and 10 for the Cramér–von Mises, Anderson–Darling and Watson kernels and for the quadrature methods introduced above. The performance of MC (illustrated by the median in solid gray line and the 2.5% and 97.5% quantiles in dashed gray line on 1000 replications) is quite poor in all cases (this can be seen also from the graphs in the Supplementary Material), in line with the bounds in Koltchinskii and Giné (2000) and the equivalence between rates of convergence for eigenvalues and eigenfunctions. The

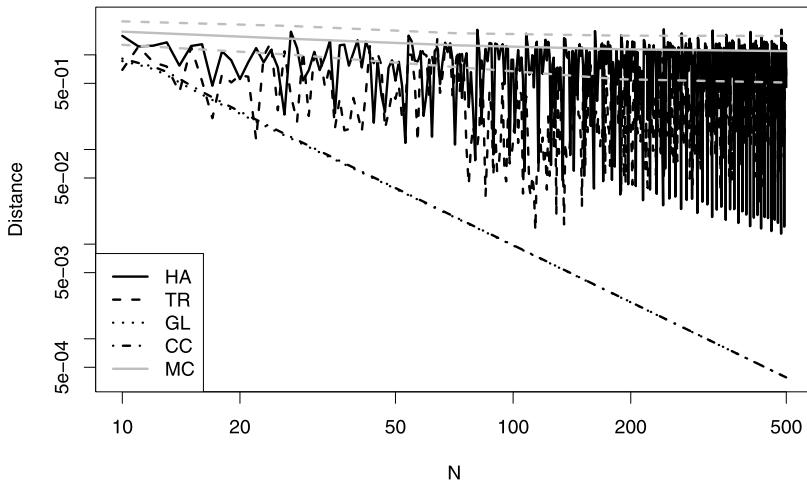


Fig. 10. Distance between the true and the approximated first 6 eigenfunctions of the Watson kernel.

Table 2

Cumulative distribution function of the Watson statistic and approximation computed through the algorithm described in the text with eigenvalues approximated through HA.

x	cdf				relative error		
	Exact	N = 10	N = 100	N = 1000	N = 10	N = 100	N = 1000
0.025	0.034001	0.067275	0.034767	0.034011	$3.44 \cdot 10^{-2}$	$7.93 \cdot 10^{-4}$	$1.02 \cdot 10^{-5}$
0.050	0.292900	0.314682	0.293363	0.292904	$3.08 \cdot 10^{-2}$	$6.55 \cdot 10^{-4}$	$6.68 \cdot 10^{-6}$
0.075	0.550283	0.550562	0.550255	0.550282	$6.20 \cdot 10^{-4}$	$-6.21 \cdot 10^{-5}$	$-1.78 \cdot 10^{-6}$
0.100	0.722922	0.715010	0.722732	0.722920	$-2.86 \cdot 10^{-2}$	$-6.87 \cdot 10^{-4}$	$-8.42 \cdot 10^{-6}$
0.125	0.830494	0.820963	0.830282	0.830491	$-5.62 \cdot 10^{-2}$	$-1.25 \cdot 10^{-3}$	$-1.42 \cdot 10^{-5}$
0.150	0.896468	0.887791	0.896283	0.896466	$-8.38 \cdot 10^{-2}$	$-1.79 \cdot 10^{-3}$	$-1.97 \cdot 10^{-5}$
0.175	0.936787	0.929702	0.936641	0.936785	$-1.12 \cdot 10^{-1}$	$-2.32 \cdot 10^{-3}$	$-2.53 \cdot 10^{-5}$
0.200	0.961408	0.955950	0.961298	0.961406	$-1.41 \cdot 10^{-1}$	$-2.84 \cdot 10^{-3}$	$-3.08 \cdot 10^{-5}$
0.225	0.976439	0.972388	0.976360	0.976438	$-1.72 \cdot 10^{-1}$	$-3.37 \cdot 10^{-3}$	$-3.65 \cdot 10^{-5}$
0.250	0.985616	0.982685	0.985560	0.985616	$-2.04 \cdot 10^{-1}$	$-3.90 \cdot 10^{-3}$	$-4.22 \cdot 10^{-5}$
0.275	0.991219	0.989137	0.991180	0.991218	$-2.37 \cdot 10^{-1}$	$-4.42 \cdot 10^{-3}$	$-4.80 \cdot 10^{-5}$
0.300	0.994639	0.993182	0.994613	0.994639	$-2.72 \cdot 10^{-1}$	$-4.95 \cdot 10^{-3}$	$-5.39 \cdot 10^{-5}$
0.325	0.996727	0.995719	0.996709	0.996727	$-3.08 \cdot 10^{-1}$	$-5.48 \cdot 10^{-3}$	$-5.99 \cdot 10^{-5}$
0.350	0.998002	0.997311	0.997990	0.998002	$-3.46 \cdot 10^{-1}$	$-6.01 \cdot 10^{-3}$	$-6.60 \cdot 10^{-5}$
0.375	0.998780	0.998310	0.998772	0.998780	$-3.85 \cdot 10^{-1}$	$-6.54 \cdot 10^{-3}$	$-7.22 \cdot 10^{-5}$
0.400	0.999255	0.998938	0.999250	0.999255	$-4.27 \cdot 10^{-1}$	$-7.07 \cdot 10^{-3}$	$-7.84 \cdot 10^{-5}$

behavior of TR, when defined, is very disparate: indeed, in the Cramér–von Mises case, TR performs very well while, for the Watson kernel, it is very bad. This seems to be due to a shift in the eigenfunctions that is very difficult to explain: the eigenfunctions are accurately described, but they are shifted of a small amount, sufficient to worsen dramatically the performance of the algorithm. This is clear from the graphs in the Supplementary Material. The same phenomenon takes place for the quasi-Monte Carlo method and worsens its performance. GL performs quite well, but the best method is CC. In particular, for the Anderson–Darling kernel, CC with $N = 10$ points is much better than HA with $N = 250$ points!

5.4. Cumulative distribution function

In order to show the performance of the Davies algorithm, the cdf of the Watson statistic at a set of points has been computed and the values have been compared with the result of the Davies algorithm using the eigenvalues as obtained from HA (Table 2), TR (Table 3), GL (Table 4) and CC (Table 5) with 10, 100 and 1000 points.

The Watson statistic has been chosen because its asymptotic distribution is related to the asymptotic distribution of the Kolmogorov–Smirnov statistic (see Watson, 1961, p. 112), as both can be expressed as the distribution of the supremum of a Brownian bridge. For the asymptotic distribution of the Kolmogorov–Smirnov statistic a rapidly convergent series representation (see Watson, 1961, p. 112, or Serfling, 1980, Theorem A, p. 62) has been used.

Table 3

Cumulative distribution function of the Watson statistic and approximation computed through the algorithm described in the text with eigenvalues approximated through TR.

x	cdf				relative error		
	Exact	N = 10	N = 100	N = 1000	N = 10	N = 100	N = 1000
0.025	0.034001	0.068770	0.034353	0.034005	$3.60 \cdot 10^{-2}$	$3.64 \cdot 10^{-4}$	$3.59 \cdot 10^{-6}$
0.050	0.292900	0.312309	0.293049	0.292901	$2.74 \cdot 10^{-2}$	$2.12 \cdot 10^{-4}$	$2.08 \cdot 10^{-6}$
0.075	0.550283	0.547966	0.550245	0.550282	$-5.15 \cdot 10^{-3}$	$-8.43 \cdot 10^{-5}$	$-8.32 \cdot 10^{-7}$
0.100	0.722922	0.713869	0.722839	0.722922	$-3.27 \cdot 10^{-2}$	$-3.03 \cdot 10^{-4}$	$-2.97 \cdot 10^{-6}$
0.125	0.830494	0.820953	0.830411	0.830493	$-5.63 \cdot 10^{-2}$	$-4.85 \cdot 10^{-4}$	$-4.77 \cdot 10^{-6}$
0.150	0.896468	0.888348	0.896400	0.896467	$-7.84 \cdot 10^{-2}$	$-6.56 \cdot 10^{-4}$	$-6.45 \cdot 10^{-6}$
0.175	0.936787	0.930448	0.936735	0.936787	$-1.00 \cdot 10^{-1}$	$-8.23 \cdot 10^{-4}$	$-8.09 \cdot 10^{-6}$
0.200	0.961408	0.956687	0.961369	0.961407	$-1.22 \cdot 10^{-1}$	$-9.90 \cdot 10^{-4}$	$-9.72 \cdot 10^{-6}$
0.225	0.976439	0.973030	0.976412	0.976439	$-1.45 \cdot 10^{-1}$	$-1.16 \cdot 10^{-3}$	$-1.13 \cdot 10^{-5}$
0.250	0.985616	0.983207	0.985597	0.985616	$-1.68 \cdot 10^{-1}$	$-1.32 \cdot 10^{-3}$	$-1.30 \cdot 10^{-5}$
0.275	0.991219	0.989543	0.991206	0.991219	$-1.91 \cdot 10^{-1}$	$-1.49 \cdot 10^{-3}$	$-1.46 \cdot 10^{-5}$
0.300	0.994639	0.993489	0.994630	0.994639	$-2.14 \cdot 10^{-1}$	$-1.65 \cdot 10^{-3}$	$-1.62 \cdot 10^{-5}$
0.325	0.996727	0.995946	0.996721	0.996727	$-2.39 \cdot 10^{-1}$	$-1.82 \cdot 10^{-3}$	$-1.79 \cdot 10^{-5}$
0.350	0.998002	0.997476	0.997998	0.998002	$-2.63 \cdot 10^{-1}$	$-1.98 \cdot 10^{-3}$	$-1.95 \cdot 10^{-5}$
0.375	0.998780	0.998428	0.998778	0.998780	$-2.89 \cdot 10^{-1}$	$-2.15 \cdot 10^{-3}$	$-2.11 \cdot 10^{-5}$
0.400	0.999255	0.999021	0.999254	0.999255	$-3.14 \cdot 10^{-1}$	$-2.32 \cdot 10^{-3}$	$-2.27 \cdot 10^{-5}$

Table 4

Cumulative distribution function of the Watson statistic and approximation computed through the algorithm described in the text with eigenvalues approximated through GL.

x	cdf				relative error		
	Exact	N = 10	N = 100	N = 1000	N = 10	N = 100	N = 1000
0.025	0.034001	0.073507	0.034562	0.034007	$4.09 \cdot 10^{-2}$	$5.80 \cdot 10^{-4}$	$5.88 \cdot 10^{-6}$
0.050	0.292900	0.317032	0.293139	0.292902	$3.41 \cdot 10^{-2}$	$3.38 \cdot 10^{-4}$	$3.40 \cdot 10^{-6}$
0.075	0.550283	0.548650	0.550223	0.550282	$-3.63 \cdot 10^{-3}$	$-1.34 \cdot 10^{-4}$	$-1.36 \cdot 10^{-6}$
0.100	0.722922	0.712554	0.722789	0.722921	$-3.74 \cdot 10^{-2}$	$-4.83 \cdot 10^{-4}$	$-4.88 \cdot 10^{-6}$
0.125	0.830494	0.819151	0.830362	0.830492	$-6.69 \cdot 10^{-2}$	$-7.75 \cdot 10^{-4}$	$-7.82 \cdot 10^{-6}$
0.150	0.896468	0.886670	0.896359	0.896467	$-9.46 \cdot 10^{-2}$	$-1.05 \cdot 10^{-3}$	$-1.06 \cdot 10^{-5}$
0.175	0.936787	0.929073	0.936704	0.936786	$-1.22 \cdot 10^{-1}$	$-1.31 \cdot 10^{-3}$	$-1.33 \cdot 10^{-5}$
0.200	0.961408	0.955630	0.961347	0.961407	$-1.50 \cdot 10^{-1}$	$-1.58 \cdot 10^{-3}$	$-1.59 \cdot 10^{-5}$
0.225	0.976439	0.972247	0.976396	0.976439	$-1.78 \cdot 10^{-1}$	$-1.84 \cdot 10^{-3}$	$-1.86 \cdot 10^{-5}$
0.250	0.985616	0.982641	0.985586	0.985616	$-2.07 \cdot 10^{-1}$	$-2.11 \cdot 10^{-3}$	$-2.13 \cdot 10^{-5}$
0.275	0.991219	0.989143	0.991198	0.991219	$-2.36 \cdot 10^{-1}$	$-2.37 \cdot 10^{-3}$	$-2.39 \cdot 10^{-5}$
0.300	0.994639	0.993209	0.994625	0.994639	$-2.67 \cdot 10^{-1}$	$-2.64 \cdot 10^{-3}$	$-2.66 \cdot 10^{-5}$
0.325	0.996727	0.995753	0.996718	0.996727	$-2.98 \cdot 10^{-1}$	$-2.91 \cdot 10^{-3}$	$-2.93 \cdot 10^{-5}$
0.350	0.998002	0.997344	0.997996	0.998002	$-3.30 \cdot 10^{-1}$	$-3.17 \cdot 10^{-3}$	$-3.19 \cdot 10^{-5}$
0.375	0.998780	0.998339	0.998776	0.998780	$-3.62 \cdot 10^{-1}$	$-3.44 \cdot 10^{-3}$	$-3.46 \cdot 10^{-5}$
0.400	0.999255	0.998961	0.999253	0.999255	$-3.95 \cdot 10^{-1}$	$-3.70 \cdot 10^{-3}$	$-3.73 \cdot 10^{-5}$

Table 5

Cumulative distribution function of the Watson statistic and approximation computed through the algorithm described in the text with eigenvalues approximated through CC.

x	cdf				relative error		
	Exact	N = 10	N = 100	N = 1000	N = 10	N = 100	N = 1000
0.025	0.034001	0.076871	0.034568	0.034007	$4.44 \cdot 10^{-2}$	$5.86 \cdot 10^{-4}$	$5.89 \cdot 10^{-6}$
0.050	0.292900	0.319590	0.293141	0.292902	$3.77 \cdot 10^{-2}$	$3.42 \cdot 10^{-4}$	$3.41 \cdot 10^{-6}$
0.075	0.550283	0.548880	0.550222	0.550282	$-3.12 \cdot 10^{-3}$	$-1.35 \cdot 10^{-4}$	$-1.37 \cdot 10^{-6}$
0.100	0.722922	0.711701	0.722787	0.722921	$-4.05 \cdot 10^{-2}$	$-4.88 \cdot 10^{-4}$	$-4.88 \cdot 10^{-6}$
0.125	0.830494	0.818048	0.830361	0.830492	$-7.34 \cdot 10^{-2}$	$-7.83 \cdot 10^{-4}$	$-7.83 \cdot 10^{-6}$
0.150	0.896468	0.885654	0.896358	0.896467	$-1.04 \cdot 10^{-1}$	$-1.06 \cdot 10^{-3}$	$-1.06 \cdot 10^{-5}$
0.175	0.936787	0.928244	0.936703	0.936786	$-1.35 \cdot 10^{-1}$	$-1.33 \cdot 10^{-3}$	$-1.33 \cdot 10^{-5}$
0.200	0.961408	0.954993	0.961346	0.961407	$-1.66 \cdot 10^{-1}$	$-1.60 \cdot 10^{-3}$	$-1.60 \cdot 10^{-5}$
0.225	0.976439	0.971775	0.976395	0.976439	$-1.98 \cdot 10^{-1}$	$-1.86 \cdot 10^{-3}$	$-1.86 \cdot 10^{-5}$
0.250	0.985616	0.982300	0.985586	0.985616	$-2.31 \cdot 10^{-1}$	$-2.13 \cdot 10^{-3}$	$-2.13 \cdot 10^{-5}$
0.275	0.991219	0.988901	0.991198	0.991219	$-2.64 \cdot 10^{-1}$	$-2.40 \cdot 10^{-3}$	$-2.40 \cdot 10^{-5}$
0.300	0.994639	0.993040	0.994625	0.994639	$-2.98 \cdot 10^{-1}$	$-2.67 \cdot 10^{-3}$	$-2.66 \cdot 10^{-5}$
0.325	0.996727	0.995636	0.996718	0.996727	$-3.33 \cdot 10^{-1}$	$-2.93 \cdot 10^{-3}$	$-2.93 \cdot 10^{-5}$
0.350	0.998002	0.997263	0.997996	0.998002	$-3.70 \cdot 10^{-1}$	$-3.20 \cdot 10^{-3}$	$-3.20 \cdot 10^{-5}$
0.375	0.998780	0.998284	0.998776	0.998780	$-4.07 \cdot 10^{-1}$	$-3.47 \cdot 10^{-3}$	$-3.46 \cdot 10^{-5}$
0.400	0.999255	0.998924	0.999253	0.999255	$-4.45 \cdot 10^{-1}$	$-3.74 \cdot 10^{-3}$	$-3.73 \cdot 10^{-5}$

Table 6

Relative error of Davies algorithm with real weights and with weights approximated by the different methods for the Watson statistic.

	$N = 10$	$N = 100$	$N = 1000$
True weights	0.1666667	0.019607840	0.001996008
HA	0.4266450	0.007066250	$7.843184 \cdot 10^{-5}$
TR	0.3141730	0.002316861	$2.273100 \cdot 10^{-5}$
GL	0.3954432	0.003700858	$3.727959 \cdot 10^{-5}$
CC	0.4450957	0.003738163	$3.731690 \cdot 10^{-5}$

The tables also display the relative error given by:

$$\frac{\mathbb{P} \left\{ \sum_{j=1}^{\infty} \lambda_j Z_j^2 \leq x \right\} - \mathbb{P} \left\{ \sum_{j=1}^N \hat{\lambda}_{N,j} Z_j^2 \leq x \right\}}{\mathbb{P} \left\{ \sum_{j=1}^{\infty} \lambda_j Z_j^2 > x \right\}}$$

for the same values of x . The function has been normalized using the survival function in order to emphasize the properties of the algorithm for the most relevant values of x .

The results show that the approximation with $N = 100$ is already good enough for most applications, while the approximation based on $N = 1000$ points is extremely good and has always at least 5 digits (4 for HA) coinciding with the real value of the cdf. The approximation is better in the tail (that is for probabilities larger than 95%), since in this case even N as small as 10 can be good enough! Moreover, the present results are similar to the ones outlined in Section 5.1, as GL and CC have asymptotically equivalent performances (so that no digit differs for $N = 1000$), and TR slightly outperforms GL and CC even if the difference appears to be almost negligible.

A further fact that is not apparent from the tables is shown in Table 6. Here the first row shows the maximum relative error when N true weights are used to approximate the cdf of the infinite weighted sum $\sum_{j=1}^{\infty} \lambda_j Z_j^2$. In practice, each cell in the first row shows the maximum value of:

$$\frac{\left| \mathbb{P} \left\{ \sum_{j=1}^{\infty} \lambda_j Z_j^2 \leq x \right\} - \mathbb{P} \left\{ \sum_{j=1}^N \lambda_j Z_j^2 \leq x \right\} \right|}{\mathbb{P} \left\{ \sum_{j=1}^{\infty} \lambda_j Z_j^2 > x \right\}}$$

over the values of x used in Tables 2, 3, 4 and 5. The other rows show the same quantity when the weights are computed using the quadrature rules presented before. It is apparent that truncation of the spectrum performs much worse than the Wielandt–Nyström method. As explained at the end of Section 4, the reason is that the approximation of $\sum_{j=1}^{\infty} \lambda_j Z_j^2$ is much better when the approximating variable retains the expectation of the approximated variable (see the bounds in Chourat and Seri, 2013; Seri, 2017). From this point of view, the Wielandt–Nyström method is better than truncation of the true spectrum.

5.5. Application to the Hall test

This section contains an application to the test presented in Hall (1985, p. 127), that is a V -statistic with kernel:

$$h(x, y) = (\rho - \|x - y\|) \cdot \mathbf{1}_{\{\|x-y\| \leq \rho\}} - \rho^2$$

for $\rho < \frac{1}{2}$ and irrational, where $\|\cdot\|$ is the distance on the interval $[0, 1]$ wrapped around a circle, namely $\|x - y\| = (y - x) \wedge (x - y + 1)$ for $0 \leq x \leq y \leq 1$. The eigenvalues are given by $\lambda_j = 2 \left(2 \left[\frac{j}{2} \right] \pi \right)^{-2} \left(1 - \cos \left(2 \left[\frac{j}{2} \right] \pi \rho \right) \right)$ (the leading 2 lacks in the formula in Hall, 1985, p. 127, because of a different definition of the spectral decomposition) and the eigenfunctions are $\phi_j(x) = 2^{\frac{1}{2}} \cos \left(2 \left[\frac{j}{2} \right] \pi x \right)$ for j even and $\phi_j(x) = 2^{\frac{1}{2}} \sin \left(2 \left[\frac{j}{2} \right] \pi x \right)$ for j odd. In the computations below, $\rho = \pi/8$ is taken.

Only the Clenshaw–Curtis method is considered in these computations since it appears to be the most performing one. Fig. 11 displays the computed eigenvalues as a function of N : eigenvalues corresponding to even and odd indexes are displayed respectively as solid and dashed lines, while the real values are represented by black dots. Eigenfunctions for $N = 10, 20, 40, 80, 160, 320$ are displayed in Fig. 12: the black curves represent the eigenfunctions for $N = 320$ while lighter curves identify lower values of N . For $N = 10$ the first eigenfunctions are correctly recovered while the last ones show some deviation from the true curves; however, the algorithm correctly describes the overall behavior of the true functions for $N = 40$. Table 7 shows the value of the cdf of the test statistic for $N = 10, 100, 1000$.

5.6. Application to the Schilling test

This section contains an application to the statistic proposed in Bickel and Breiman (1983) and Schilling (1983a,b). Let $(\mathbf{X}_1, \dots, \mathbf{X}_n)$ be a sample from a bounded density in \mathbb{R}^m continuous on an open support. Let g be the hypothesized density,

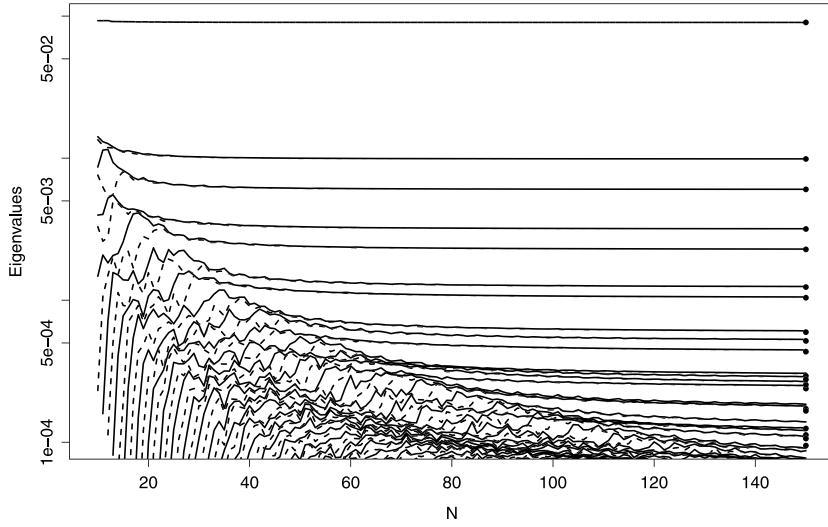


Fig. 11. Eigenvalues of the Hall kernel evaluated through CC as a function of N .

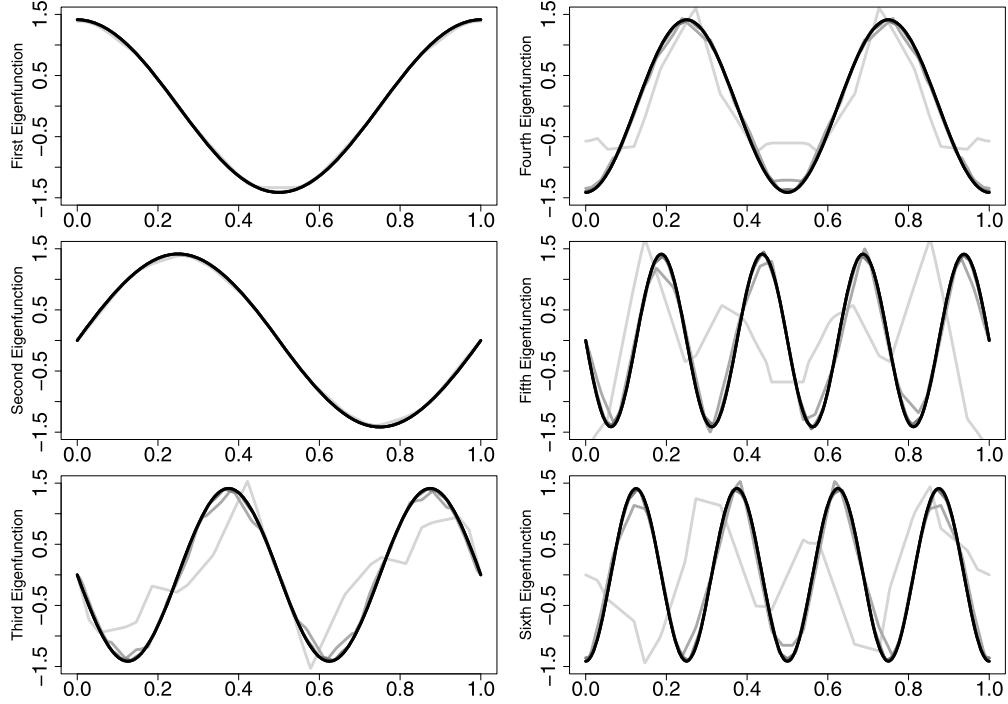


Fig. 12. Performance of the algorithm for the approximation of the eigenfunctions of the Hall kernel evaluated through CC: eigenfunction ϕ_j for $j = 1, 2, 3$ from top to bottom in the first column and for $j = 4, 5, 6$ from top to bottom in the second column; approximated eigenfunctions for $N = 10, 20, 40, 80, 160$ displayed in increasing shades of grey, true eigenfunction (equal to the approximated one with $N = 320$) in black.

$R_i := \min_{j \neq i} \|\mathbf{X}_j - \mathbf{X}_i\|$ the distance from \mathbf{X}_i to its nearest neighbor, and $V(r)$ the volume of an m -sphere of radius r . Then, define:

$$W_i := \exp \{-ng(\mathbf{X}_i) V(R_i)\}, \quad i = 1, \dots, n.$$

If g is the true density, then the W_i 's have an asymptotically uniform distribution. For a bounded continuous weight function $w(x)$, one can consider the weighted empirical process:

$$F_n(t) = \frac{1}{n} \sum_{i=1}^n w(\mathbf{X}_i) \mathbf{1}\{W_i \leq t\}, \quad 0 \leq t \leq 1,$$

Table 7

Cumulative distribution function of the Hall statistic and approximation computed through the algorithm described in the text with eigenvalues approximated through CC.

x	cdf				relative error		
	Exact	N = 10	N = 100	N = 1000	N = 10	N = 100	N = 1000
0.125	0.305492	0.316818	0.305590	0.305493	$1.63 \cdot 10^{-2}$	$1.41 \cdot 10^{-4}$	$1.43 \cdot 10^{-6}$
0.250	0.651396	0.648297	0.651356	0.651396	$-8.89 \cdot 10^{-3}$	$-1.16 \cdot 10^{-4}$	$-1.17 \cdot 10^{-6}$
0.375	0.825604	0.820590	0.825547	0.825603	$-2.88 \cdot 10^{-2}$	$-3.25 \cdot 10^{-4}$	$-3.30 \cdot 10^{-6}$
0.500	0.912756	0.908498	0.912709	0.912755	$-4.88 \cdot 10^{-2}$	$-5.34 \cdot 10^{-4}$	$-5.43 \cdot 10^{-6}$
0.625	0.956355	0.953332	0.956322	0.956355	$-6.93 \cdot 10^{-2}$	$-7.43 \cdot 10^{-4}$	$-7.55 \cdot 10^{-6}$
0.750	0.978166	0.976199	0.978145	0.978166	$-9.01 \cdot 10^{-2}$	$-9.52 \cdot 10^{-4}$	$-9.68 \cdot 10^{-6}$
0.875	0.989077	0.987861	0.989064	0.989077	$-1.11 \cdot 10^{-1}$	$-1.16 \cdot 10^{-3}$	$-1.18 \cdot 10^{-5}$
1.000	0.994536	0.993809	0.994528	0.994536	$-1.33 \cdot 10^{-1}$	$-1.37 \cdot 10^{-3}$	$-1.39 \cdot 10^{-5}$
1.125	0.997266	0.996842	0.997262	0.997266	$-1.55 \cdot 10^{-1}$	$-1.58 \cdot 10^{-3}$	$-1.61 \cdot 10^{-5}$
1.250	0.998632	0.998390	0.998630	0.998632	$-1.78 \cdot 10^{-1}$	$-1.79 \cdot 10^{-3}$	$-1.82 \cdot 10^{-5}$
1.375	0.999316	0.999179	0.999315	0.999316	$-2.01 \cdot 10^{-1}$	$-2.00 \cdot 10^{-3}$	$-2.03 \cdot 10^{-5}$
1.500	0.999658	0.999581	0.999657	0.999658	$-2.24 \cdot 10^{-1}$	$-2.21 \cdot 10^{-3}$	$-2.24 \cdot 10^{-5}$
1.625	0.999829	0.999786	0.999828	0.999829	$-2.48 \cdot 10^{-1}$	$-2.42 \cdot 10^{-3}$	$-2.46 \cdot 10^{-5}$
1.750	0.999914	0.999891	0.999914	0.999914	$-2.72 \cdot 10^{-1}$	$-2.63 \cdot 10^{-3}$	$-2.67 \cdot 10^{-5}$
1.875	0.999957	0.999944	0.999957	0.999957	$-2.97 \cdot 10^{-1}$	$-2.83 \cdot 10^{-3}$	$-2.88 \cdot 10^{-5}$
2.000	0.999979	0.999972	0.999978	0.999979	$-3.22 \cdot 10^{-1}$	$-3.04 \cdot 10^{-3}$	$-3.09 \cdot 10^{-5}$

and:

$$Z_n(t) := \sqrt{n} \{F_n(t) - \mathbb{E}_g F_n(t)\}, \quad 0 \leq t \leq 1,$$

where \mathbb{E}_g is the expectation under g . It can be shown that, if g is the true density, the process $Z_n(t)$ converges weakly to a Gaussian process with mean zero and covariance kernel:

$$k(x, y) = \left[x(1 + y \ln x) + xy \int_{B(x,y)} \{\eta(x, y, \omega) - 1\} d\omega \right] \mathbb{E}_g w^2(\mathbf{X}_1)$$

$$- xy(1 + \ln xy + \ln x \ln y) (\mathbb{E}_g w(\mathbf{X}_1))^2, \quad 0 \leq x \leq y \leq 1,$$

where:

$$\mathbf{B}(x, y) = \{\boldsymbol{\omega} \in \mathbb{R}^m : r(x) \leq \|\boldsymbol{\omega}\| \leq r(x) + r(y)\},$$

$$\ln \eta(x, y, \boldsymbol{\omega}) = \int_{\{\mathbf{z} \in \mathbb{R}^m : \|\mathbf{z}\| \leq r(x), \|\mathbf{z} - \boldsymbol{\omega}\| \leq r(y)\}} d\mathbf{z},$$

and $r(\cdot)$ is the radius of an m -sphere with volume $-\ln(\cdot)$. The integral can be solved explicitly only in the case $m = 1$, and approximated in the case $m \rightarrow \infty$. A natural test is to reject the null if $\int_0^1 Z_n^2(t) dt$ is large. It can be shown (see Schilling, 1983a, Section 3) that $\int_0^1 Z_n^2(t) dt$ converges in distribution to a weighted sum of χ^2 random variables whose weights are the eigenvalues of the integral operator induced by the kernel $k(\cdot, \cdot)$. Note that here $k(\cdot, \cdot)$ is not the kernel of the V -statistic, but the covariance function of the Gaussian process. Nevertheless, its eigenvalues can still be recovered through the same algorithm.

In the case $m = 1$:

$$k(x, y) = \left[x(1 + y \ln x) + xy \left(\ln y + 2y^{-\frac{1}{2}} - 2 \right) \right] \mathbb{E}_g w^2(\mathbf{X}_1)$$

$$- xy(1 + \ln xy + \ln x \ln y) (\mathbb{E}_g w(\mathbf{X}_1))^2, \quad 0 \leq x \leq y \leq 1.$$

In the case $m = \infty$, one gets:

$$k(x, y) = [x(1 + y \ln x) + xy \ln x \ln y] \mathbb{E}_g w^2(\mathbf{X}_1)$$

$$- xy(1 + \ln xy + \ln x \ln y) (\mathbb{E}_g w(\mathbf{X}_1))^2, \quad 0 \leq x \leq y \leq 1.$$

Two interesting cases arise, respectively, when $w(\cdot) \equiv 1$ or when $\mathbb{E}_g w(\mathbf{X}_1) = 0$ and $\mathbb{E}_g w^2(\mathbf{X}_1) = 1$. Schilling (1983b) computed the distribution of the four versions of the statistic with the following covariance functions:

Table 8

Cumulative distribution function of the Schilling statistic with $m = 1$ and $w(\cdot) \equiv 1$ computed through the algorithm described in the text with eigenvalues approximated through CC.

x	$N = 10$	$N = 100$	$N = 1000$	Schilling
0.05	0.06233527	0.03136490	0.03098597	0.034
0.10	0.25203374	0.22997143	0.22975772	0.236
0.15	0.43921421	0.43437447	0.43434537	0.439
0.20	0.58390528	0.58679351	0.58683385	0.590
0.25	0.68955400	0.69507872	0.69513965	0.697
0.30	0.76617700	0.77222678	0.77228989	0.764
0.35	0.82219143	0.82794881	0.82800744	0.829
0.40	0.86361884	0.86880020	0.86885227	0.870
0.45	0.89461802	0.89915937	0.89920464	0.900
0.50	0.91806083	0.92198661	0.92202554	0.923
0.60	0.94971654	0.95259078	0.95261908	0.9529
0.70	0.96867436	0.97074994	0.97077026	0.9709
0.80	0.98027011	0.98175678	0.98177126	0.9819
0.90	0.98747150	0.98852963	0.98853988	0.9886
1.00	0.99199413	0.99274316	0.99275038	0.9928
1.10	0.99485857	0.99538623	0.99539129	0.9954
1.20	0.99668473	0.99705487	0.99705841	0.9971
1.30	0.99785506	0.99811375	0.99811621	0.9981
1.40	0.99860831	0.99878852	0.99879022	0.9988
1.50	0.99909484	0.99922001	0.99922119	0.9992
1.60	0.99941004	0.99949677	0.99949758	—
1.70	0.99961476	0.99967472	0.99967528	—
1.75	0.99968851	0.99973832	0.99973879	0.9998
1.80	0.99974804	0.99978941	0.99978979	—
1.90	0.99983496	0.99986346	0.99986372	—
2.00	0.99989176	0.99991136	0.99991153	0.99991
2.10	0.99992892	0.99994238	0.99994250	—
2.25	0.99996211	0.99996975	0.99996982	0.99997
2.50	0.99998666	0.99998961	0.99998964	—
3.00	0.99999832	0.99999876	0.99999876	—
4.00	0.99999997	0.99999998	0.99999998	—
5.00	1.00000000	1.00000000	1.00000000	—

Table 9

Cumulative distribution function of the Schilling statistic with $m = 1$, $\mathbb{E}_g w(X_1) = 0$ and $\mathbb{E}_g w^2(X_1) = 1$ computed through the algorithm described in the text with eigenvalues approximated through CC.

x	$N = 10$	$N = 100$	$N = 1000$	Schilling
0.05	0.02222570	0.00980884	0.00966330	0.011
0.10	0.11317478	0.09759467	0.09743398	0.101
0.15	0.22808151	0.21892039	0.21883489	0.222
0.20	0.33709072	0.33279698	0.33275967	0.335
0.25	0.43215096	0.43074542	0.43073518	0.433
0.30	0.51291222	0.51318054	0.51318571	0.515
0.35	0.58106505	0.58230866	0.58232274	0.584
0.40	0.63858735	0.64039149	0.64041063	0.642
0.45	0.68726656	0.68937431	0.68939612	0.691
0.50	0.72860686	0.73085446	0.73087740	0.732
0.60	0.79399309	0.79624153	0.79626414	0.797
0.70	0.84217224	0.84424379	0.84426445	0.845
0.80	0.87813373	0.87996941	0.87998763	0.880
0.90	0.90527943	0.90687351	0.90688928	0.907
1.00	0.92596942	0.92733852	0.92735202	0.928
1.10	0.94186935	0.94303790	0.94304940	0.943
1.20	0.95417382	0.95516752	0.95517729	0.9553
1.30	0.96375238	0.96459542	0.96460370	0.9647
1.40	0.97124644	0.97196047	0.97196748	0.9721
1.50	0.97713470	0.97773868	0.97774460	0.9778
1.60	0.98177813	0.98228845	0.98229344	0.9823
1.70	0.98545138	0.98588210	0.98588631	0.9859
1.75	0.98699202	0.98738758	0.98739144	0.9874
1.80	0.98836501	0.98872819	0.98873174	0.9888
1.90	0.99068154	0.99098747	0.99099046	0.9910
2.00	0.99252714	0.99278460	0.99278712	0.9928
2.10	0.99400020	0.99421668	0.99421879	0.9942
2.25	0.99567530	0.99584193	0.99584355	0.9959
2.50	0.99748282	0.99759010	0.99759114	0.9976
3.00	0.99913608	0.99917996	0.99918038	0.9992
4.00	0.99989460	0.99990166	0.99990173	0.99990
5.00	0.99998676	0.99998785	0.99998786	0.99999

Table 10

Cumulative distribution function of the Schilling statistic with $m = \infty$ and $w(\cdot) \equiv 1$ computed through the algorithm described in the text with eigenvalues approximated through CC.

x	$N = 10$	$N = 100$	$N = 1000$	Schilling
0.05	0.05719199	0.03290107	0.03261189	0.036
0.10	0.22387932	0.20776500	0.20761142	0.212
0.15	0.38713618	0.38249155	0.38245620	0.386
0.20	0.51631440	0.51679229	0.51680323	0.519
0.25	0.61456999	0.61703195	0.61705961	0.619
0.30	0.68951260	0.69264827	0.69268109	0.694
0.35	0.74743146	0.75069611	0.75072942	0.752
0.40	0.79285016	0.79601705	0.79604898	0.797
0.45	0.82894447	0.83192283	0.83195267	0.833
0.50	0.85795951	0.86071804	0.86074557	0.861
0.60	0.90077638	0.90309156	0.90311458	0.904
0.70	0.92978564	0.93169955	0.93171853	0.932
0.80	0.94984916	0.95141580	0.95143131	0.9516
0.90	0.96392896	0.96520095	0.96521352	0.9653
1.00	0.97391503	0.97494053	0.97495064	0.9750
1.10	0.98105461	0.98187632	0.98188440	0.9819
1.20	0.98619106	0.98684600	0.98685243	0.9869
1.30	0.98990492	0.99042458	0.99042967	0.9905
1.40	0.99260124	0.99301195	0.99301596	0.9930
1.50	0.99456553	0.99488904	0.99489219	0.9949
1.60	0.99600070	0.99625478	0.99625725	0.9963
1.70	0.99705191	0.99725095	0.99725288	0.9973
1.75	0.99746740	0.99764342	0.99764512	—
1.80	0.99782357	0.99797914	0.99798065	0.9980
1.90	0.99839110	0.99851246	0.99851363	0.9985
2.00	0.99880921	0.99890371	0.99890462	0.9989
2.10	0.99911770	0.99919117	0.99919188	0.9992
2.25	0.99943626	0.99948649	0.99948698	0.9995
2.50	0.99973166	0.99975816	0.99975841	0.9998
3.00	0.99993843	0.99994568	0.99994575	0.99995
4.00	0.99999664	0.99999716	0.99999717	—
5.00	0.99999981	0.99999985	0.99999985	—

Table 11

Cumulative distribution function of the Schilling statistic with $m = \infty$, $\mathbb{E}_g w(X_1) = 0$ and $\mathbb{E}_g w^2(X_1) = 1$ computed through the algorithm described in the text with eigenvalues approximated through CC.

x	$N = 10$	$N = 100$	$N = 1000$	Schilling
0.05	0.02038592	0.01044095	0.01032602	0.011
0.10	0.10047678	0.08881489	0.08869694	0.091
0.15	0.20072691	0.19324671	0.19317566	0.196
0.20	0.29726078	0.29289793	0.29285735	0.295
0.25	0.38369072	0.38131005	0.38128826	0.383
0.30	0.45938848	0.45831036	0.45830092	0.460
0.35	0.52526975	0.52508702	0.52508614	0.526
0.40	0.58254644	0.58299433	0.58299950	0.584
0.45	0.63237982	0.63327292	0.63328235	0.634
0.50	0.67579733	0.67699969	0.67701207	0.678
0.60	0.74680297	0.74834274	0.74835831	0.749
0.70	0.80122495	0.80286467	0.80288111	0.803
0.80	0.84320234	0.84480930	0.84482533	0.845
0.90	0.87577514	0.87727949	0.87729445	0.878
1.00	0.90119155	0.90256096	0.90257454	0.903
1.10	0.92112627	0.92235030	0.92236242	0.923
1.20	0.93683568	0.93791629	0.93792697	0.938
1.30	0.94926902	0.95021474	0.95022408	0.9504
1.40	0.95914828	0.95997078	0.95997889	0.9601
1.50	0.96702619	0.96773823	0.96774524	0.9678
1.60	0.97332853	0.97394278	0.97394882	0.9740
1.70	0.97838512	0.97891360	0.97891879	0.9790
1.75	0.98052891	0.98101869	0.98102351	—
1.80	0.98245292	0.98290663	0.98291109	0.9830
1.90	0.98573306	0.98612190	0.98612572	0.9862
2.00	0.98838374	0.98871651	0.98871978	0.9887
2.10	0.99052989	0.99081433	0.99081712	0.9908
2.25	0.99301421	0.99323856	0.99324075	0.9933
2.50	0.99577223	0.99592263	0.99592410	0.9959
3.00	0.99842919	0.99849594	0.99849659	0.9985
4.00	0.99977476	0.99978744	0.99978756	0.99997
5.00	0.99996668	0.99996901	0.99996903	0.99997

$$k(x, y) = \begin{cases} x + xy \left(2y^{-\frac{1}{2}} - 3 - \ln x \ln y \right), & m = 1, w(\cdot) \equiv 1, \\ x + xy \left(\ln xy + 2y^{-\frac{1}{2}} - 2 \right), & m = 1, \mathbb{E}_g w(\mathbf{X}_1) = 0, \mathbb{E}_g w^2(\mathbf{X}_1) = 1, \\ x - xy (1 + \ln y), & m = \infty, w(\cdot) \equiv 1, \\ x + xy \ln x (1 + \ln y), & m = \infty, \mathbb{E}_g w(\mathbf{X}_1) = 0, \mathbb{E}_g w^2(\mathbf{X}_1) = 1, \end{cases}$$

for $0 \leq x \leq y \leq 1$; for $0 \leq y \leq x \leq 1$, the covariance functions are defined by symmetry. His algorithm for retrieving the eigenvalues is complex. It consists of two steps: first, the Wielandt–Nyström method with $N = 100$, nodes $(\frac{1}{2N}, \frac{3}{2N}, \dots, 1 - \frac{1}{2N})$ and equal weights is used; second, the approximated eigenvalues are used as the starting point of another algorithm approximating the eigenfunctions and the kernel appearing in the integral equation through power series. The author states that the difference between the eigenvalues obtained in the first and in the second step is negligible. The computational results are shown in Tables 8, 9, 10 and 11. The last column contains the values obtained in Schilling (1983b).

6. Conclusions

An algorithm, based on the replacement of an integral with a quadrature rule, has been proposed for the computation of the eigenvalues and eigenfunctions of an integral operator associated with the kernel of a quadratic statistic. Its computational properties have been reviewed for several choices of the quadrature rule and of the kernel. The quantities can be used for the computation of the asymptotic distribution of the statistic, and of its asymptotic power and efficiency. The Clenshaw–Curtis quadrature rule seems to be the most reliable, and provides results that are consistently among the best for both the eigenvalues and the eigenfunctions.

Appendix A. Auxiliary results and proofs

A.1. The Davies algorithm

As explained in Section 3, the approximation of the eigenvalues of the integral operator \mathcal{H} through the finite spectrum of the matrix \mathbf{H}_N makes it possible to replace the weighted (infinite) sum of χ^2 random variables in (3) with a quadratic form in Gaussian random variables.

Some methods for the approximation of the distribution of this class of random variables have been proposed in the literature, starting from the seminal paper of Imhof (1961). This distribution can be computed through the techniques of Imhof (1961), Sheil and O’Muircheartaigh (1977) and Davies (1973, 1980) for quadratic forms in normal random variables. See Mathai and Provost (1992) for an overview of some of these algorithms. Related algorithms have been proposed in Rice (1980), Brown (1986) and Lindsay et al. (2000).

Here, the algorithm in Davies (1980) is used. The difference between Davies’s and Imhof’s algorithms is the fact that the latter does not provide control for the so-called integration error (see below for more details). Imhof (1961, p. 423) states explicitly that “[i]t does not seem feasible [...] to obtain an upper bound for the error of integration resulting from the application of a standard quadrature formula.” However, the former uses a Fourier cosine series summation formula to approximate the integral through a trapezium quadrature rule and provide a control on the error. With respect to the algorithm of Sheil and O’Muircheartaigh, the one by Davies holds more generally because the former cannot handle negative weights. Indeed, negative eigenvalues can arise for kernels that are not positive semidefinite or, even if the kernel is positive semidefinite, when the version of the Wielandt–Nyström method recommended in Koltchinskii and Giné (2000) is used.

The algorithm can be briefly described as follows. Let X be a random variable with characteristic function $\phi(u) = \mathbb{E}(e^{iuX})$. Under some conditions (i.e. if $\mathbb{E}|X| < \infty$ and, for some c and $\delta > 0$ and for all $u > 1$, $|\phi(u)| < cu^{-\delta}$; see Gil-Pelaez, 1951, and Davies, 1973, p. 415), it is possible to express the cdf of X as a function of ϕ :

$$\mathbb{P}(X < x) = \frac{1}{2} - \int_{-\infty}^{+\infty} \Im \left(\frac{\phi(u) e^{-iux}}{2\pi u} \right) du$$

where $\Im(\cdot)$ is the imaginary part of a complex number. In Davies (1973), it is shown that:

$$\begin{aligned} \mathbb{P}(X < x) &= \frac{1}{2} - \sum_{k=0}^{\infty} \frac{\Im \left(\phi \left((k + \frac{1}{2}) \Delta \right) e^{-i(k + \frac{1}{2}) \Delta x} \right)}{\pi (k + \frac{1}{2})} \\ &\quad - \sum_{n=1}^{\infty} (-1)^n \left\{ \mathbb{P} \left(X < x - \frac{2\pi n}{\Delta} \right) - \mathbb{P} \left(X > x + \frac{2\pi n}{\Delta} \right) \right\}. \end{aligned}$$

As a result, the author proposes to approximate $\mathbb{P}(X < x)$ as:

$$\mathbb{P}(X < x) \simeq \frac{1}{2} - \sum_{k=0}^K \frac{\Im\left(\phi((k + \frac{1}{2})\Delta)e^{-i(k + \frac{1}{2})\Delta x}\right)}{\pi(k + \frac{1}{2})}.$$

Using the characteristic formula for a finite sum of χ^2 random variables, i.e. $\phi(u) = \prod_{j=1}^N (1 - 2i\lambda_j u)^{-\frac{1}{2}}$, this formula becomes:

$$\mathbb{P}(X < x) \simeq \frac{1}{2} - \sum_{k=0}^K \frac{\sin\left(\sum_{j=1}^N \frac{\arctan[2(k + \frac{1}{2})\Delta\lambda_j]}{2} - (k + \frac{1}{2})\Delta x\right)}{\pi(k + \frac{1}{2}) \prod_{j=1}^N \left(1 + 4(k + \frac{1}{2})^2 \Delta^2 \lambda_j^2\right)^{\frac{1}{4}}}.$$

The use of this formula introduces two sources of error:

- an *integration error* $-\sum_{n=1}^{\infty} (-1)^n \left\{ \mathbb{P}\left(X < x - \frac{2\pi n}{\Delta}\right) - \mathbb{P}\left(X > x + \frac{2\pi n}{\Delta}\right) \right\}$; this error can be made small choosing adequately Δ . Davies (1973, p. 416) proposes to choose Δ so that $\max\{\mathbb{P}\left(X < x - \frac{2\pi}{\Delta}\right), \mathbb{P}\left(X > x + \frac{2\pi}{\Delta}\right)\}$ is less than half the maximum allowable error.
- a *truncation error* $-\sum_{k=K+1}^{\infty} \frac{\Im\left(\phi((k + \frac{1}{2})\Delta)e^{-i(k + \frac{1}{2})\Delta x}\right)}{\pi(k + \frac{1}{2})}$: the way in which this error term can be made small choosing adequately the parameter K is discussed in Imhof (1961, p. 423) and more thoroughly in Davies (1980, p. 324) where several upper bounds on this error are given.

The last source of error comes from the replacement of the true eigenvalues $(\lambda_j)_j$ with the computed ones $(\hat{\lambda}_{N,j})_{j=1,\dots,N}$. If the weights $(\lambda_j)_j$ and $(\hat{\lambda}_{N,j})_j$ are non-negative, Theorem 1 in Naumov et al. (2018) yields:

$$\begin{aligned} & \sup_{x \geq 0} \left| \mathbb{P}\left(\sum_{j=1}^N \hat{\lambda}_{N,j} Z_j^2 \leq x\right) - \mathbb{P}\left(\sum_{j=1}^{\infty} \lambda_j Z_j^2 \leq x\right) \right| \\ & \leq C \left\{ \left(\sum_{j=1}^{\infty} \lambda_j^2 \sum_{j=2}^{\infty} \lambda_j^2 \right)^{-1/4} + \left(\sum_{j=1}^N \hat{\lambda}_{N,j}^2 \sum_{j=2}^N \hat{\lambda}_{N,j}^2 \right)^{-1/4} \right\} \\ & \quad \cdot \left(\sum_{j=1}^N |\lambda_j - \hat{\lambda}_{N,j}| + \sum_{k=N+1}^{\infty} |\lambda_j| \right) \\ & \lesssim C \left(\sum_{j=2}^{\infty} \lambda_j^2 \right)^{-\frac{1}{2}} \left(\sum_{j=1}^N |\lambda_j - \hat{\lambda}_{N,j}| + \sum_{k=N+1}^{\infty} |\lambda_j| \right) \end{aligned}$$

for an absolute constant $C > 0$ that can differ from place to place. When the eigenvalues $\hat{\lambda}_{N,j}$ for $j = 1, \dots, N$ are known exactly, the bound gives the error in the truncation of $\sum_{j=1}^{\infty} \lambda_j Z_j^2$ to $\sum_{j=1}^N \lambda_j Z_j^2$. However, in that case computable and tight bounds can be proved along the lines of Choirat and Seri (2013) and Seri (2017).

A.2. Proofs

Proof of Corollary 1. It is straightforward to note that $\alpha_N, \beta_N, \gamma_N$ and ρ_N can be majorized by $\|\eta_N\|_{\infty}$. Therefore, $\lambda_j^2 \vee \hat{\lambda}_{N,j}^2 - \rho_N \geq \lambda_j^2 - \|\eta_N\|_{\infty}$ and, provided $\lambda_j^2 - \|\eta_N\|_{\infty} > 0$:

$$\begin{aligned} |\hat{\lambda}_{N,1} - \lambda_1| & \leq \frac{\gamma_N}{(\lambda_1^2 \vee \hat{\lambda}_{N,1}^2 - \rho_N)^{\frac{1}{2}}} \leq \frac{\|\eta_N\|_{\infty}}{(\lambda_1^2 - \|\eta_N\|_{\infty})^{\frac{1}{2}}}, \\ |\hat{\lambda}_{N,j} - \lambda_j| & \leq \frac{\gamma_N + \rho_N}{(\lambda_j^2 \vee \hat{\lambda}_{N,j}^2 - \rho_N)^{\frac{1}{2}}} \leq \frac{2\|\eta_N\|_{\infty}}{(\lambda_j^2 - \|\eta_N\|_{\infty})^{\frac{1}{2}}}. \end{aligned}$$

The bound:

$$\left| \hat{\lambda}_{N,j} - \lambda_j \right| \leq \left(\frac{1+\sqrt{5}}{2} (\gamma_N + \rho_N) \right)^{\frac{1}{2}} \leq \left((1+\sqrt{5}) \|\eta_N\|_\infty \right)^{\frac{1}{2}}$$

does not require any condition. The final result is obtained comparing the bounds. QED

Proof of Proposition 1. From (6), the formula for the Nyström extension is:

$$\hat{\phi}_{N,j}(x) = \frac{1}{\hat{\lambda}_{N,j}} \sum_{k=1}^N h(x, y_k) w_k [\tilde{\phi}_{N,j}]_k.$$

For $x \in [y_\ell, y_{\ell+1})$, the behavior of the approximated eigenfunction can be obtained setting $x = y_\ell + \varepsilon$ with $\varepsilon \in [0, y_{\ell+1} - y_\ell)$. Then:

$$\hat{\phi}_{N,j}(y_\ell + \varepsilon) - \hat{\phi}_{N,j}(y_\ell) = \frac{1}{\hat{\lambda}_{N,j}} \sum_{k=1}^N \{h(y_\ell + \varepsilon, y_k) - h(y_\ell, y_k)\} w_k [\tilde{\phi}_{N,j}]_k.$$

For the Cramér–von Mises kernel:

$$h(y_\ell + \varepsilon, y_k) - h(y_\ell, y_k) = \varepsilon y_\ell + \frac{\varepsilon^2}{2} + y_\ell \vee y_k - (y_\ell + \varepsilon) \vee y_k.$$

Using the fact that $\sum_{k=1}^N w_k [\tilde{\phi}_{N,j}]_k = 0$:

$$\begin{aligned} \hat{\phi}_{N,j}(y_\ell + \varepsilon) - \hat{\phi}_{N,j}(y_\ell) &= \frac{1}{\hat{\lambda}_{N,j}} \sum_{k=1}^N \{y_\ell \vee y_k - (y_\ell + \varepsilon) \vee y_k\} w_k [\tilde{\phi}_{N,j}]_k \\ &= -\frac{\varepsilon}{\hat{\lambda}_{N,j}} \sum_{k=1}^\ell w_k [\tilde{\phi}_{N,j}]_k. \end{aligned}$$

The same takes place for the Watson kernel. Indeed:

$$h(y_\ell + \varepsilon, y_k) - h(y_\ell, y_k) = \frac{\varepsilon^2}{2} + (y_\ell - y_k) \varepsilon + \frac{\varepsilon}{2} + y_\ell \vee y_k - (y_\ell + \varepsilon) \vee y_k$$

and:

$$\begin{aligned} \hat{\phi}_{N,j}(y_\ell + \varepsilon) - \hat{\phi}_{N,j}(y_\ell) &= \frac{1}{\hat{\lambda}_{N,j}} \sum_{k=1}^N \{-y_k \varepsilon + y_\ell \vee y_k - (y_\ell + \varepsilon) \vee y_k\} w_k [\tilde{\phi}_{N,j}]_k \\ &= -\frac{\varepsilon}{\hat{\lambda}_{N,j}} \sum_{k=1}^N y_k w_k [\tilde{\phi}_{N,j}]_k - \frac{\varepsilon}{\hat{\lambda}_{N,j}} \sum_{k=1}^\ell w_k [\tilde{\phi}_{N,j}]_k. \end{aligned}$$

This does not happen for the Anderson–Darling kernel. Indeed:

$$h(y_\ell + \varepsilon, y_k) - h(y_\ell, y_k) = -\ln((y_\ell + \varepsilon) \vee y_k - (y_\ell + \varepsilon) y_k) + \ln(y_\ell \vee y_k - y_\ell y_k)$$

and:

$$\begin{aligned} \hat{\phi}_{N,j}(y_\ell + \varepsilon) - \hat{\phi}_{N,j}(y_\ell) &= \frac{1}{\hat{\lambda}_{N,j}} \sum_{k=1}^\ell \{-\ln((y_\ell + \varepsilon) - (y_\ell + \varepsilon) y_k) + \ln(y_\ell - y_\ell y_k)\} w_k [\tilde{\phi}_{N,j}]_k \\ &\quad + \frac{1}{\hat{\lambda}_{N,j}} \sum_{k=\ell+1}^N \{-\ln(y_k - (y_\ell + \varepsilon) y_k) + \ln(y_k - y_\ell y_k)\} w_k [\tilde{\phi}_{N,j}]_k \\ &= \frac{-\ln(y_\ell + \varepsilon) + \ln y_\ell}{\hat{\lambda}_{N,j}} \sum_{k=1}^\ell w_k [\tilde{\phi}_{N,j}]_k \\ &\quad + \frac{-\ln(1 - y_\ell - \varepsilon) + \ln(1 - y_\ell)}{\hat{\lambda}_{N,j}} \sum_{k=\ell+1}^N w_k [\tilde{\phi}_{N,j}]_k \end{aligned}$$

$$= \ln \left(\frac{y_\ell (1 - y_\ell - \varepsilon)}{(1 - y_\ell) (y_\ell + \varepsilon)} \right) \frac{1}{\hat{\lambda}_{N,j}} \sum_{k=1}^{\ell} w_k [\tilde{\phi}_{N,j}]_k$$

where the last step uses the equality $\sum_{k=\ell+1}^N w_k [\tilde{\phi}_{N,j}]_k = -\sum_{k=1}^{\ell} w_k [\tilde{\phi}_{N,j}]_k$, coming from $\sum_{k=1}^N w_k [\tilde{\phi}_{N,j}]_k = 0$. QED

A.3. Rate of decrease for the MC case

In this section, the results of Section 4 in Koltchinskii and Giné (2000) are used to investigate the rate of convergence of the approximated spectrum for the MC case. However, while their result is stated for a matrix $\check{\mathbf{H}}_N$ whose generic element is $\check{h}_{ik,N} = \frac{1-\delta_{ik}}{N} h(y_i, y_k)$ where δ_{ik} is the Kronecker delta, the present paper uses the matrix \mathbf{H}_N whose generic element is $h_{ik,N} = \frac{1}{N} h(y_i, y_k)$ (the situation is further complicated by the fact that they call H_n the first matrix and \check{H}_n the second one). This can be solved by showing that the distance between the spectra of $\check{\mathbf{H}}_N$ and \mathbf{H}_N is negligible with respect to the distance between the spectra of \mathbf{H}_N and \mathcal{H} . Let $\check{\lambda}_{N,j}$ be the j -th eigenvalue of $\check{\mathbf{H}}_N$. Eq. (3.10) in Koltchinskii and Giné (2000) bounds the L^2 -distance of the spectra of $\check{\mathbf{H}}_N$ and \mathbf{H}_N as:

$$\sum_{j=1}^N (\check{\lambda}_{N,j} - \hat{\lambda}_{N,j})^2 \leq \sum_{i=1}^N \left(\frac{1}{N} \sum_{j=1}^{\infty} \lambda_j \phi_j^2(y_i) \right)^2 = \frac{1}{N^2} \sum_{i=1}^N h^2(y_i, y_i).$$

In most cases of interest (e.g., if the kernel h is bounded, or if the nodes are well chosen), the right-hand side is $O(N^{-1})$ and $\sqrt{\sum_{j=1}^N (\check{\lambda}_{N,j} - \hat{\lambda}_{N,j})^2} = O(N^{-\frac{1}{2}})$.

However, when the points are iid, a solution under less stringent assumptions can be obtained noting that:

$$\mathbb{E} \sum_{j=1}^N (\check{\lambda}_{N,j} - \hat{\lambda}_{N,j})^2 \leq \frac{1}{N} \int_0^1 h^2(x, x) dx = O(N^{-1})$$

for all kernels such that $\int_0^1 h^2(x, x) dx < \infty$, and $\sqrt{\sum_{j=1}^N (\check{\lambda}_{N,j} - \hat{\lambda}_{N,j})^2} = O_{\mathbb{P}}(N^{-\frac{1}{2}})$. This is always of a smaller order than the bounds established in the following.

A.3.1. Cramér–von Mises statistic

In this case $\phi_j(x) = \sqrt{2} \cos(\pi jx)$ and:

$$\begin{aligned} \int_0^1 \phi_k^2(x) \phi_j^2(x) dx &= 4 \int_0^1 \cos^2(\pi kx) \cos^2(\pi jx) dx \\ &= \int_0^1 [\cos(2\pi kx) + 1][\cos(2\pi jx) + 1] dx \\ &= 1 + \frac{1}{2} \delta_{jk} \leq \frac{3}{2}. \end{aligned}$$

Therefore, using $\lambda_j = j^{-2}\pi^{-2}$ in Theorem 4.2 in Koltchinskii and Giné (2000), $\mathbb{E}(\Lambda^{(2)}(N))^2 = O(\frac{R}{N} + R^{-3})$ and, under the optimal choice $R \sim N^{\frac{1}{4}}$, $\mathbb{E}(\Lambda^{(2)}(N))^2 = O(R^{-\frac{3}{4}})$ and $\Lambda^{(2)}(N) = O_{\mathbb{P}}(N^{-\frac{3}{8}})$.

A.3.2. Watson statistic

In this case $\phi_j(x) = \sqrt{2} \sin(\pi(j+1)x)$ for j odd and $\phi_j(x) = \sqrt{2} \cos(\pi jx)$ for j even:

$$\int_0^1 \phi_k^2(x) \phi_j^2(x) dx = \begin{cases} 4 \int_0^1 \sin^2(\pi(k+1)x) \sin^2(\pi(j+1)x) dx, & k \text{ odd, } j \text{ odd}, \\ 4 \int_0^1 \sin^2(\pi(k+1)x) \cos^2(\pi jx) dx, & k \text{ odd, } j \text{ even}, \\ 4 \int_0^1 \cos^2(\pi kx) \sin^2(\pi(j+1)x) dx, & k \text{ even, } j \text{ odd}, \\ 4 \int_0^1 \cos^2(\pi kx) \cos^2(\pi jx) dx, & k \text{ even, } j \text{ even}, \end{cases}$$

$$= \begin{cases} 1 + \frac{1}{2}\delta_{jk}, & k \text{ odd, } j \text{ odd,} \\ 1 - \frac{1}{2}\delta_{j,k+1}, & k \text{ odd, } j \text{ even,} \\ 1 - \frac{1}{2}\delta_{j+1,k}, & k \text{ even, } j \text{ odd,} \\ 1 + \frac{1}{2}\delta_{jk}, & k \text{ even, } j \text{ even.} \end{cases}$$

As a result, $\int_0^1 \phi_k^2(x) \phi_j^2(x) dx \leq \frac{3}{2}$. Theorem 4.2 in Koltchinskii and Giné (2000) yields the same rate as for the Cramér-von Mises case.

A.3.3. Anderson–Darling statistic

From Cauchy–Schwarz inequality:

$$\int_0^1 \phi_k^2(x) \phi_j^2(x) dx \leq \sqrt{\int_0^1 \phi_k^4(x) dx \int_0^1 \phi_j^4(x) dx}.$$

It is clear that $\int_0^1 \phi_k^4(x) dx = \frac{(2k+1)^2}{2} \int_{-1}^1 P_k^4(x) dx$. Now, from 34.3.19 in Olver et al. (2010):

$$P_k(x) P_\ell(x) = \sum_{m=|k-\ell|}^{k+\ell} \begin{pmatrix} k & \ell & m \\ 0 & 0 & 0 \end{pmatrix}^2 (2m+1) P_m(x)$$

where $\begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & m_3 \end{pmatrix}$ is the Wigner 3j symbol (see Olver et al., 2010, Chapter 34). Therefore:

$$P_k^4(x) = \sum_{m=0}^{2k} \begin{pmatrix} k & k & m \\ 0 & 0 & 0 \end{pmatrix}^4 (2m+1)^2 P_m^2(x)$$

and:

$$\begin{aligned} \int_{-1}^1 P_k^4(x) dx &= \sum_{m=0}^{2k} \begin{pmatrix} k & k & m \\ 0 & 0 & 0 \end{pmatrix}^4 (2m+1)^2 \int_{-1}^1 P_m^2(x) dx \\ &= 2 \sum_{m=0}^{2k} \begin{pmatrix} k & k & m \\ 0 & 0 & 0 \end{pmatrix}^4 (2m+1) = 2 \sum_{n=0}^k \begin{pmatrix} k & k & 2n \\ 0 & 0 & 0 \end{pmatrix}^4 (4n+1) \end{aligned}$$

where the second equality derives from the fact that, when the projective quantum numbers (the lower parameters) are all zero, the Wigner 3j symbol is non-zero only when the sum of the angular momenta (the upper parameters) is even. Now, from 34.3.5 in Olver et al. (2010):

$$\begin{pmatrix} k & k & 2n \\ 0 & 0 & 0 \end{pmatrix}^4 = \left[\frac{(2k-2n)!}{(2k+2n+1)!} \right]^2 \left[\frac{(k+n)!}{(k-n)!} \frac{(2n)!}{(2n)!} \right]^4 \frac{1}{(n!)^8}.$$

It is easy to see that

$$\begin{pmatrix} k & k & 0 \\ 0 & 0 & 0 \end{pmatrix}^4 = \frac{1}{(2k+1)^2} \sim \frac{1}{4k^2},$$

and, through Stirling formula:

$$\begin{pmatrix} k & k & 2k \\ 0 & 0 & 0 \end{pmatrix}^4 = \left[\frac{1}{(4k+1)!} \right]^2 \left[\frac{(2k)!}{k!} \right]^8 \sim \frac{1}{8\pi k^3}.$$

For all values of n with $0 < n < k$, let a be defined by the equality $n = ak$. Stirling formula yields:

$$\begin{aligned} \begin{pmatrix} k & k & 2ak \\ 0 & 0 & 0 \end{pmatrix}^4 &= \left(\frac{(2k(1-a))!}{(2k(1+a)+1)!} \right)^2 \left\{ \frac{(2ak)((1+a)k)!}{[(ak)!]^2 ((1-a)k)!} \right\}^4 \\ &\sim \frac{1}{4\pi^2 a^2 (1-a^2) k^4} \sim \frac{1}{4\pi^2 n^2 (k^2 - n^2)}. \end{aligned}$$

As a result:

$$\begin{aligned} & \sum_{n=0}^k \binom{k \ k \ 2n}{0 \ 0 \ 0}^4 (4n+1) \\ & \sim \frac{1}{4k^2} + \frac{4k}{8\pi k^3} + \frac{1}{\pi^2} \sum_{n=1}^{k-1} \frac{1}{n(k^2 - n^2)} \\ & = \frac{1}{4k^2} + \frac{1}{2\pi^2 k^2} + \frac{\psi^{(0)}(k+1) + 3\psi^{(0)}(k) - \psi^{(0)}(2k) + 3\gamma}{2\pi^2 k^2} \\ & \sim \frac{3 \ln k}{2\pi^2 k^2} \end{aligned}$$

and $\int_0^1 \phi_k^4(x) dx \sim \frac{6 \ln k}{\pi^2}$. Therefore:

$$\begin{aligned} & \sum_{k,j=1}^R (\lambda_k^2 + \lambda_j^2 + \lambda_k \lambda_j) \int_0^1 \phi_k^2(x) \phi_j^2(x) dx \\ & \leq \sum_{k,j=1}^R (\lambda_k^2 + \lambda_j^2 + \lambda_k \lambda_j) \sqrt{\int_0^1 \phi_k^4(x) dx \int_0^1 \phi_j^4(x) dx} \\ & = 2 \sum_{k=1}^R \lambda_k^2 \sqrt{\int_0^1 \phi_k^4(x) dx} \sum_{j=1}^R \sqrt{\int_0^1 \phi_j^4(x) dx} + \left(\sum_{k=1}^R \lambda_k \sqrt{\int_0^1 \phi_k^4(x) dx} \right)^2 \\ & = O(R \sqrt{\ln R}). \end{aligned}$$

From Theorem 4.2 in Koltchinskii and Giné (2000):

$$\mathbb{E}(\Lambda^{(2)}(N))^2 = O(N^{-1} R \sqrt{\ln R} + R^{-3}).$$

By equating the two terms, $R^8 \ln R \asymp N^2$. Using the Lambert W function:

$$R^8 \asymp \exp W(N^2) \asymp \exp(\ln N^2 - \ln \ln N^2) \asymp \frac{N^2}{\ln N}$$

or $R \asymp \frac{N^{\frac{1}{4}}}{\ln^{\frac{1}{8}} N}$ from which $\mathbb{E}(\Lambda^{(2)}(N))^2 = O(N^{-\frac{3}{4}} \ln^{\frac{3}{8}} N)$ and $\Lambda^{(2)}(N) = O_{\mathbb{P}}(N^{-\frac{3}{8}} \ln^{\frac{3}{16}} N)$.

Appendix B. Supplementary material

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.csda.2022.107437>.

References

- Adamczak, R., Bednorz, W., 2015. Some remarks on MCMC estimation of spectra of integral operators. Bernoulli 21, 2073–2092. <https://doi.org/10.3150/14-BEJ635>.
- Alman, J., Vassilevska Williams, V., 2021. A refined laser method and faster matrix multiplication. In: Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms. SODA. SIAM, Philadelphia, PA, pp. 522–539.
- Anselone, P.M., 1967. Collectively Compact Operator Approximations. Technical Report 76. Computer Science Department, Stanford University, Stanford, CA.
- Anselone, P.M., 1971. Collectively Compact Operator Approximation Theory and Applications to Integral Equations. Prentice-Hall Series in Automatic Computation. Prentice-Hall, Inc., Englewood Cliffs, NJ.
- Atkinson, K.E., 1967a. The numerical solution of Fredholm integral equations of the second kind. SIAM J. Numer. Anal. 4, 337–348. <https://doi.org/10.1137/0704029>.
- Atkinson, K.E., 1967b. The numerical solutions of the eigenvalue problem for compact integral operators. Trans. Am. Math. Soc. 129, 458–465. <https://doi.org/10.2307/1994601>.
- Atkinson, K.E., 1976. An automatic program for linear Fredholm integral equations of the second kind. ACM Trans. Math. Softw. 2, 154–171. <https://doi.org/10.1145/355681.355686>.
- Atkinson, K.E., Han, W., 2005. Theoretical Numerical Analysis, second ed. Texts in Applied Mathematics, vol. 39. Springer, New York, NY.
- Bickel, P.J., Breiman, L., 1983. Sums of functions of nearest neighbor distances, moment bounds, limit theorems and a goodness of fit test. Ann. Probab. 11, 185–214. <https://doi.org/10.1214/aop/1176993668>.
- Borwein, J.M., Borwein, P.B., 1988. On the complexity of familiar functions and numbers. SIAM Rev. 30, 589–601. <https://doi.org/10.1137/1030134>.

- Brown, R.H., 1986. The distribution function of positive definite quadratic forms in normal random variables. *SIAM J. Sci. Stat. Comput.* 7, 689–695. <https://doi.org/10.1137/0907046>.
- Bückner, H., 1950. Konvergenzuntersuchungen bei einem algebraischen Verfahren zur näherungsweisen Lösung von Integralgleichungen. *Math. Nachr.* 3, 358–372. <https://doi.org/10.1002/mana.19490030604>.
- Bückner, H., 1952. Die Praktische Behandlung von Integral-Gleichungen. *Ergebnisse der Angewandten Mathematik*, vol. 1. Springer-Verlag, Berlin; Göttingen; Heidelberg.
- Chen, X., White, H., 1998. Central limit and functional central limit theorems for Hilbert-valued dependent heterogeneous arrays with applications. *Econom. Theory* 14, 260–284. <https://doi.org/10.1017/S0266466698142056>.
- Choirat, C., Seri, R., 2006. The asymptotic distribution of quadratic discrepancies. In: *Monte Carlo and Quasi-Monte Carlo Methods 2004*. Springer, Berlin, pp. 61–76.
- Choirat, C., Seri, R., 2013. Computational aspects of Cui–Freeden statistics for equidistribution on the sphere. *Math. Comput.* 82, 2137–2156. <https://doi.org/10.1090/S0025-5718-2013-02698-1>.
- Colombi, R., 2020. Selection tests for possibly misspecified hierarchical multinomial marginal models. *Econom. Stat.* 16, 136–147. <https://doi.org/10.1016/j.ecosta.2019.06.002>.
- Csörgő, S., 1986. Testing for normality in arbitrary dimension. *Ann. Stat.* 14, 708–723. <https://doi.org/10.1214/aos/1176349948>.
- Davies, R.B., 1973. Numerical inversion of a characteristic function. *Biometrika* 60, 415–417. <https://doi.org/10.1093/biomet/60.2.415>.
- Davies, R.B., 1980. Statistical algorithms: algorithm AS 155: the distribution of a linear combination of χ^2 random variables. *J. R. Stat. Soc., Ser. C* 29, 323–333. <https://doi.org/10.2307/2346911>.
- Delves, L.M., Mohamed, J.L., 1985. *Computational Methods for Integral Equations*. Cambridge University Press, Cambridge.
- Demmel, J., Dumitriu, I., Holtz, O., 2007. Fast linear algebra is stable. *Numer. Math.* 108, 59–91. <https://doi.org/10.1007/s00211-007-0114-x>.
- Denker, M., 1982. Statistical decision procedures and ergodic theory. In: *Ergodic Theory and Related Topics*. Vitte, 1981. In: *Mathematical Research*, vol. 12. Akademie-Verlag, Berlin, pp. 35–47.
- Dewan, I., Prakasa Rao, B.L.S., 2001. Asymptotic normality for U -statistics of associated random variables. *J. Stat. Plan. Inference* 97, 201–225. [https://doi.org/10.1016/S0378-3758\(00\)00226-3](https://doi.org/10.1016/S0378-3758(00)00226-3).
- Dewan, I., Prakasa Rao, B.L.S., 2002. Central limit theorem for U -statistics of associated random variables. *Stat. Probab. Lett.* 57, 9–15. [https://doi.org/10.1016/S0167-7152\(01\)00194-8](https://doi.org/10.1016/S0167-7152(01)00194-8).
- Duchesne, P., Lafaye De Micheaux, P., 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. <https://doi.org/10.1016/j.csda.2009.11.025>.
- Dunford, N., Schwartz, J.T., 1963. *Linear Operators. Part II: Spectral Theory. Self Adjoint Operators in Hilbert Space*. Pure and Applied Mathematics, vol. VII. Interscience Publishers John Wiley & Sons, New York, NY; London.
- Durbin, J., Knott, M., 1972. Components of Cramér–von Mises statistics. I. *J. R. Stat. Soc., Ser. B* 34, 290–307. <https://doi.org/10.1111/j.2517-6161.1972.tb00908.x>.
- Durbin, J., Knott, M., Taylor, C.C., 1975. Components of Cramér–von Mises statistics. II. *J. R. Stat. Soc., Ser. B* 37, 216–237. <https://doi.org/10.1111/j.2517-6161.1975.tb01537x>.
- Durbin, J., Knott, M., Taylor, C.C., 1977. Corrigendum: “Components of Cramér–von Mises statistics. II” (*J. R. Statist. Soc. Ser. B* 37 (1975), 216–237). *J. R. Stat. Soc., Ser. B* 39, 394. <https://doi.org/10.1111/j.2517-6161.1977.tb01642.x>.
- Fejér, L., 1933. Mechanische Quadraturen mit positiven Cotesschen Zahlen. *Math. Z.* 37, 287–309. <https://doi.org/10.1007/BF01474575>.
- Filippova, A.A., 1962. Mises’ theorem on the asymptotic behavior of functionals of empirical distribution functions and its statistical applications. *Theory Probab. Appl.* 7, 24–57. <https://doi.org/10.1137/1107003>.
- Gil-Pelaez, J., 1951. Note on the inversion theorem. *Biometrika* 38, 481–482. <https://doi.org/10.1093/biomet/38.3-4.481>.
- Gregory, G.G., 1977. Large sample theory for U -statistics and tests of fit. *Ann. Stat.* 5, 110–123. <https://doi.org/10.1214/aos/1176343744>.
- Gregory, G.G., 1980. On efficiency and optimality of quadratic tests. *Ann. Stat.* 8, 116–131. <https://doi.org/10.1214/aos/1176344895>.
- Hall, P., 1985. Tailor-made tests of goodness of fit. *J. R. Stat. Soc., Ser. B* 47, 125–131. <https://doi.org/10.1111/j.2517-6161.1985.tb01339.x>.
- Hilbert, D., 1904. *Grundzüge einer allgemeinen Theorie der linearen Integralgleichungen. (Erste Mitteilung)*. Nachrichten von der Königl. Gesellschaft der Wissenschaften zu Göttingen. Mathematisch-Physikalische Klasse 1904, pp. 49–91.
- Hilbert, D., 1912. *Grundzüge Einer Allgemeinen Theorie der Linearen Integralgleichungen. Fortschritte der Mathematischen Wissenschaften*. Druck und Verlag von B.G. Teubner, Leipzig; Berlin.
- Imhof, J.P., 1961. Computing the distribution of quadratic forms in normal variables. *Biometrika* 48, 419–426. <https://doi.org/10.1093/biomet/48.3-4.419>.
- Keller, H.B., 1965. On the accuracy of finite difference approximations to the eigenvalues of differential and integral operators. *Numer. Math.* 7, 412–419. <https://doi.org/10.1007/BF01436255>.
- Koltchinskii, V., Giné, E., 2000. Random matrix approximation of spectra of integral operators. *Bernoulli* 6, 113–167.
- Lindsay, B.G., Pilla, R.S., Basak, P., 2000. Moment-based approximations of distributions using mixtures: theory and applications. *Ann. Inst. Stat. Math.* 52, 215–230. <https://doi.org/10.1023/A:1004105603806>.
- Linz, P., 1979. *Theoretical Numerical Analysis. An Introduction to Advanced Techniques*. Wiley-Interscience [John Wiley & Sons], New York, NY.
- Mathai, A.M., Provost, S.B., 1992. *Quadratic Forms in Random Variables. Statistics: Textbooks and Monographs*, vol. 126. Marcel Dekker, Inc., New York, NY.
- Naumov, A.A., Spokoiny, V.G., Tavyrikov, Y.E., Ulyanov, V.V., 2018. Nonasymptotic estimates for the closeness of Gaussian measures on balls. *Dokl. Math.* 98, 490–493. <https://doi.org/10.1134/S1064562418060248>.
- Neyman, J., 1937. “Smooth test” for goodness of fit. *Skand. Aktuariedtskr.* 20, 149–199. <https://doi.org/10.1080/03461238.1937.10404821>.
- Niederreiter, H., 1992. *Random Number Generation and Quasi-Monte Carlo Methods*. CBMS-NSF Regional Conference Series in Applied Mathematics, vol. 63. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA.
- Olver, F.W.J., Lozier, D.W., Boisvert, R.F., Clark, C.W. (Eds.), 2010. *NIST Handbook of Mathematical Functions*. U.S. Department of Commerce. National Institute of Standards and Technology, Washington, DC. Cambridge University Press, Cambridge.
- Pan, V.Y., Chen, Z.Q., 1999. The complexity of the matrix eigenproblem. In: *Annual ACM Symposium on Theory of Computing*. Atlanta, GA, 1999. ACM, New York, NY, pp. 507–516.
- Parr, W.C., Schucany, W.R., 1982. Minimum distance estimation and components of goodness-of-fit statistics. *J. R. Stat. Soc., Ser. B* 44, 178–189. <https://doi.org/10.1111/j.2517-6161.1982.tb01197.x>.
- Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P., 2002. *Numerical Recipes in C++*. Cambridge University Press, Cambridge.
- Rakotchi, E., 1975. Numerical solution for eigenvalues and eigenfunctions of a Hermitian kernel and an error estimate. *Math. Comput.* 29, 794–805. <https://doi.org/10.1090/S0025-5718-1975-0373356-X>.
- Rice, S.O., 1980. Distribution of quadratic forms in normal random variables—evaluation by numerical integration. *SIAM J. Sci. Stat. Comput.* 1, 438–448. <https://doi.org/10.1137/0901032>.
- Robin, J.M., Smith, R.J., 2000. Tests of rank. *Econom. Theory* 16, 151–175. <https://doi.org/10.1017/S0266466600162012>.
- Rubin, H., Vitale, R.A., 1980. Asymptotic distribution of symmetric statistics. *Ann. Stat.* 8, 165–170. <https://doi.org/10.1214/aos/1176344898>.
- Schilling, M.F., 1983a. Goodness of fit testing in R^m based on the weighted empirical distribution of certain nearest neighbor statistics. *Ann. Stat.* 11, 1–12. <https://doi.org/10.1214/aos/1176346051>.

- Schilling, M.F., 1983b. An infinite-dimensional approximation for nearest neighbor goodness of fit tests. *Ann. Stat.* 11, 13–24. <https://doi.org/10.1214/aos/1176346052>.
- Schoenfeld, D.A., 1977. Asymptotic properties of tests based on linear combinations of the orthogonal components of the Cramér–von Mises statistic. *Ann. Stat.* 5, 1017–1026. <https://doi.org/10.1214/aos/1176343956>.
- Serfling, R.J., 1980. *Approximation Theorems of Mathematical Statistics*. John Wiley & Sons, New York, NY.
- Seri, R., 2015. A non-recursive formula for the higher derivatives of the Hurwitz zeta function. *J. Math. Anal. Appl.* 424, 826–834. <https://doi.org/10.1016/j.jmaa.2014.08.012>.
- Seri, R., 2017. Statistical properties of b -adic diaphonies. *Math. Comput.* 86, 799–828. <https://doi.org/10.1090/mcom/3148>.
- Seri, R., Choirat, C., 2012. Comparison of quadrature rules for the Wielandt–Nyström method with statistical applications. In: *Numerical Analysis and Applied Mathematics ICNAAM 2012: International Conference of Numerical Analysis and Applied Mathematics*. Kos, Greece, pp. 2282–2285.
- Seri, R., Martinoli, M., 2021. Asymptotic properties of the plug-in estimator of the discrete entropy under dependence. *IEEE Trans. Inf. Theory* 67, 7659–7683. <https://doi.org/10.1109/TIT.2021.3109307>.
- Sheil, J., O'Muircheartaigh, I., 1977. Statistical algorithms: algorithm AS 106: the distribution of non-negative quadratic forms in normal variables. *J. R. Stat. Soc., Ser. C* 26, 92–98. <https://doi.org/10.2307/2346884>.
- Spence, A., 1975. On the convergence of the Nyström method for the integral equation eigenvalue problem. *Numer. Math.* 25, 57–66. <https://doi.org/10.1007/BF01419528>.
- Spence, A., 1977. Error bounds and estimates for eigenvalues of integral equations. *Numer. Math.* 29, 133–147. <https://doi.org/10.1007/BF01390333>.
- Spence, A., 1979. Error bounds and estimates for eigenvalues of integral equations. II. *Numer. Math.* 32, 139–146. <https://doi.org/10.1007/BF01404870>.
- Thomas, K.S., 1974. On the approximate solution of operator equations. *Numer. Math.* 23, 231–239. <https://doi.org/10.1007/BF01400306>.
- Trefethen, L.N., 2008. Is Gauss quadrature better than Clenshaw–Curtis? *SIAM Rev.* 50, 67–87. <https://doi.org/10.1137/060659831>.
- van der Vaart, A.W., 1998. *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge.
- von Mises, R., 1947. On the asymptotic distribution of differentiable statistical functions. *Ann. Math. Stat.* 18, 309–348. <https://doi.org/10.1214/aoms/117730385>.
- Vuong, Q.H., 1989. Likelihood ratio tests for model selection and nonnested hypotheses. *Econometrica* 57, 307–333. <https://doi.org/10.2307/1912557>.
- Waldbogel, J., 2006. Fast construction of the Fejér and Clenshaw–Curtis quadrature rules. *BIT Numer. Math.* 46, 195–202. <https://doi.org/10.1007/s10543-006-0045-4>.
- Watson, G.S., 1961. Goodness-of-fit tests on a circle. *Biometrika* 48, 109–114. <https://doi.org/10.1093/biomet/48.1-2.109>.
- Wielandt, H., 1956. Error bounds for eigenvalues of symmetric integral equations. In: *Proceedings of Symposia in Applied Mathematics*. Vol. VI. *Numerical Analysis*. McGraw-Hill Book Company, Inc., New York, NY, for the American Mathematical Society, Providence, RI, pp. 261–282.
- Xiang, S., Bornemann, F., 2012. On the convergence rates of Gauss and Clenshaw–Curtis quadrature for functions of limited regularity. *SIAM J. Numer. Anal.* 50, 2581–2587. <https://doi.org/10.1137/120869845>.

Supplementary Material to “Computing the Asymptotic Distribution of 2-nd order U - and V -statistics”

Raffaello Seri¹

^a*Center for Nonlinear and Complex Systems, Università degli Studi dell’Insubria, Via Valleggio 11, 22100 Como, Italy, and DiECO, Università degli Studi dell’Insubria, Via Monte Generoso 71, 21100 Varese, Italy*

1. Graphs

1.1. Cramér-von Mises Kernel

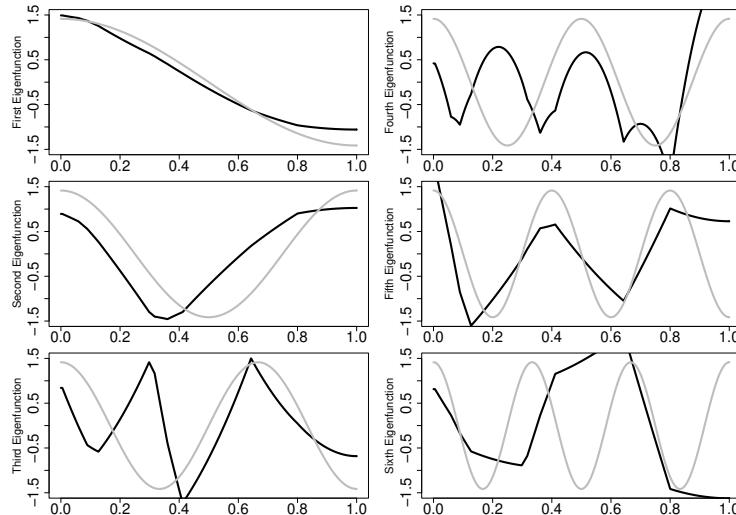


Figure 1: Eigenfunctions of the Cramér-von Mises kernel based on a Monte Carlo method (MC) with 10 points.

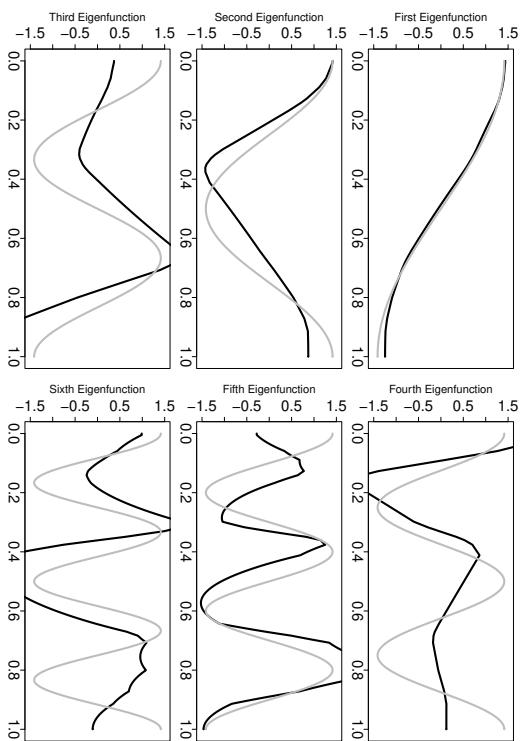


Figure 2: Eigenfunctions of the Cramér-von Mises kernel based on a Monte Carlo method (MC) with 20 points.

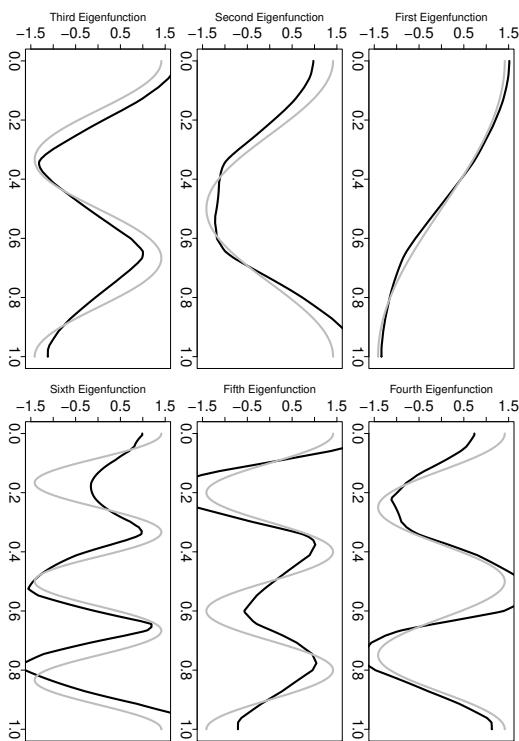


Figure 3: Eigenfunctions of the Cramér-von Mises kernel based on a Monte Carlo method (MC) with 40 points.

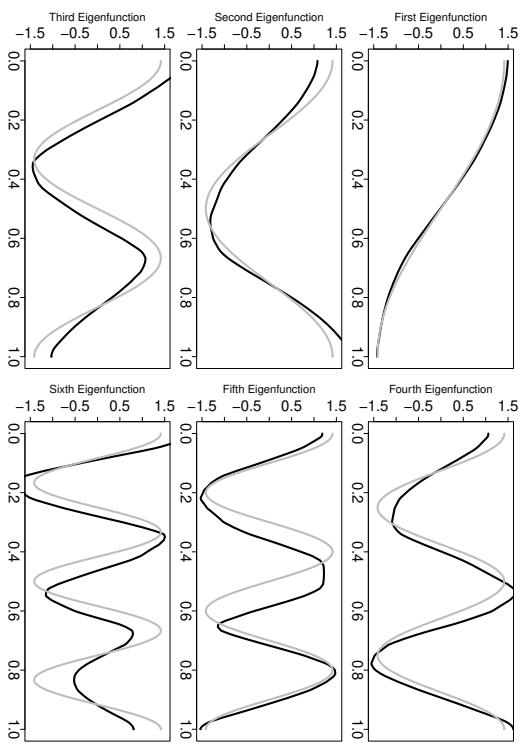


Figure 4: Eigenfunctions of the Cramér-von Mises kernel based on a Monte Carlo method (MC) with 80 points.

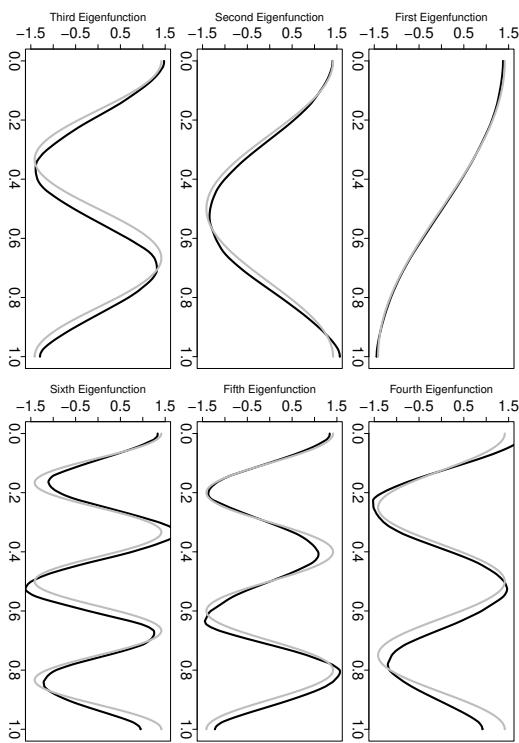


Figure 5: Eigenfunctions of the Cramér-von Mises kernel based on a Monte Carlo method (MC) with 160 points.

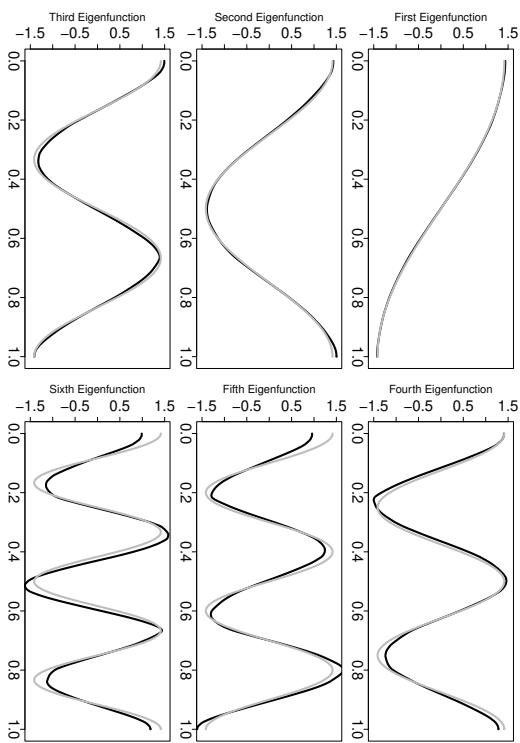


Figure 6: Eigenfunctions of the Cramér-von Mises kernel based on a Monte Carlo method (MC) with 320 points.

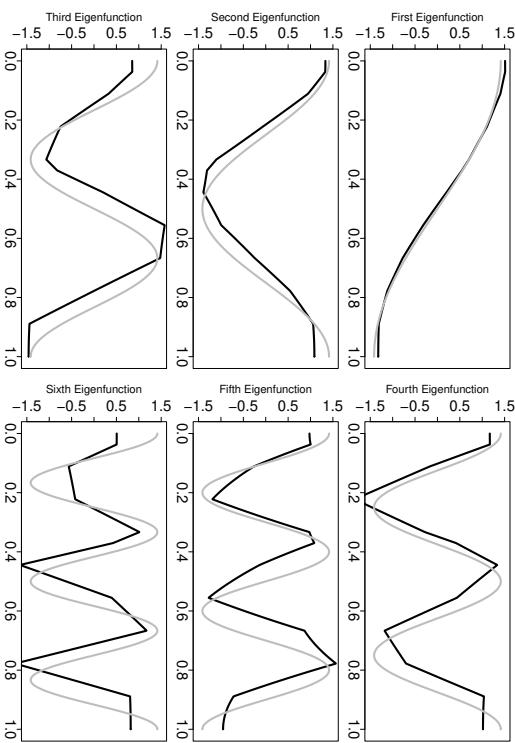


Figure 7: Eigenfunctions of the Cramér-von Mises kernel based on a quasi-Monte Carlo method (HA) with 10 points.

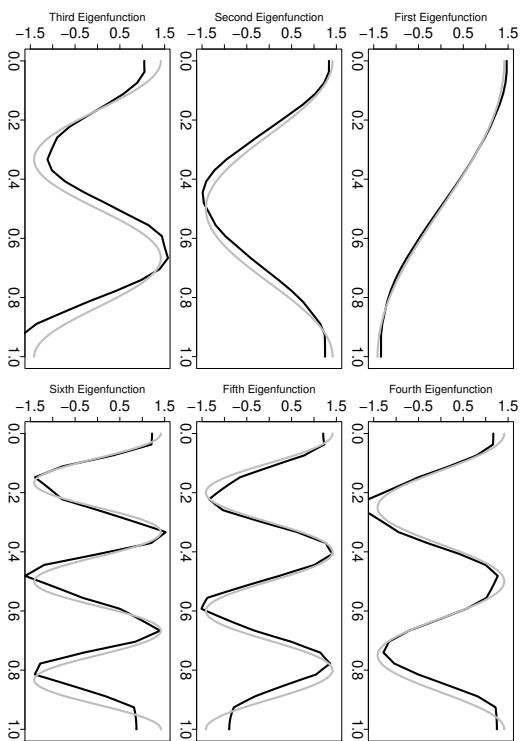


Figure 8: Eigenfunctions of the Cramér-von Mises kernel based on a quasi-Monte Carlo method (HA) with 20 points.

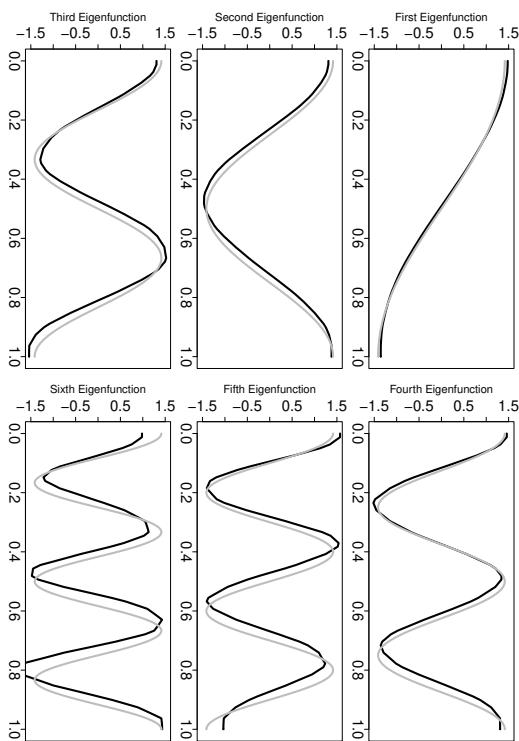


Figure 9: Eigenfunctions of the Cramér-von Mises kernel based on a quasi-Monte Carlo method (HA) with 40 points.

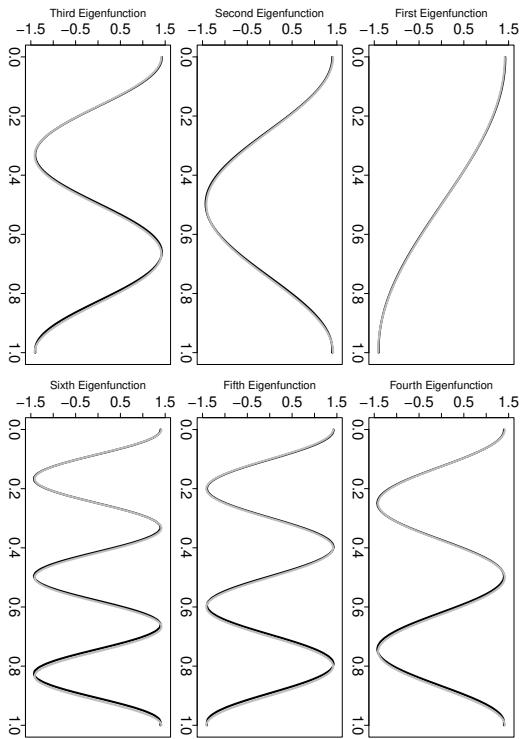


Figure 11: Eigenfunctions of the Cramér-von Mises kernel based on a quasi-Monte Carlo method (HA) with 160 points.

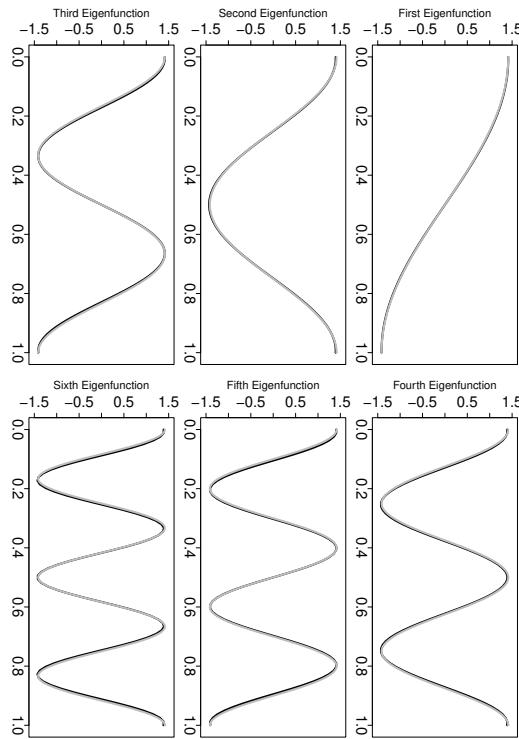


Figure 10: Eigenfunctions of the Cramér-von Mises kernel based on a quasi-Monte Carlo method (HA) with 80 points.

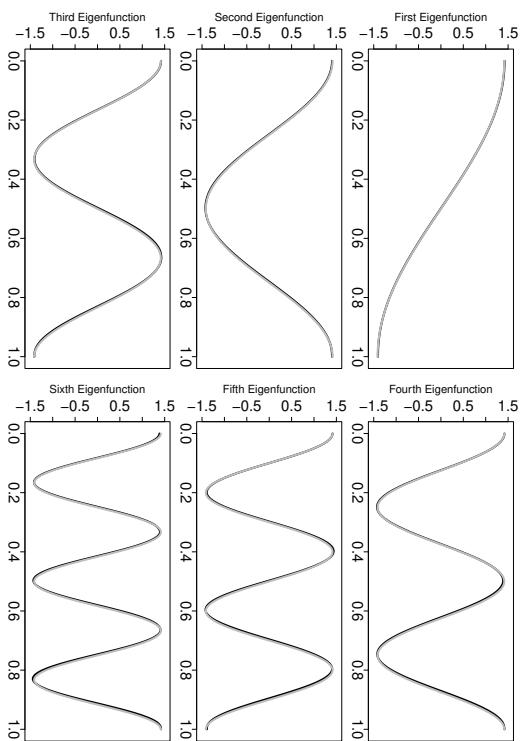


Figure 12: Eigenfunctions of the Cramér-von Mises kernel based on a quasi-Monte Carlo method (HA) with 320 points.

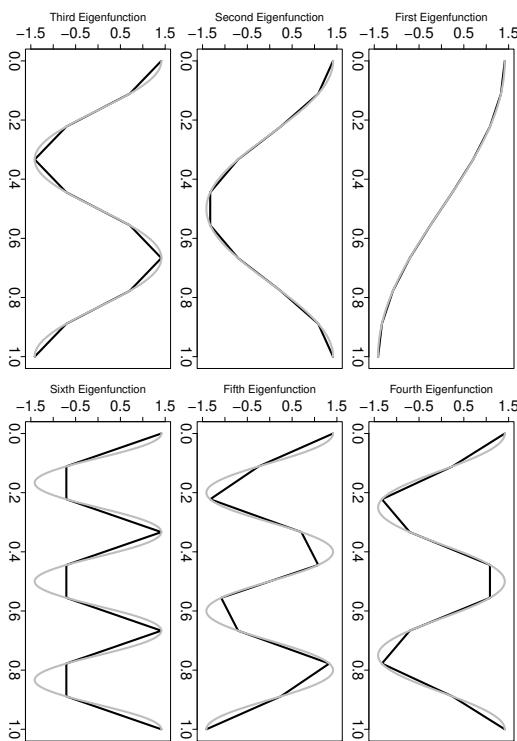


Figure 13: Eigenfunctions of the Cramér-von Mises kernel based on a trapezium rule (TR) with 10 points.

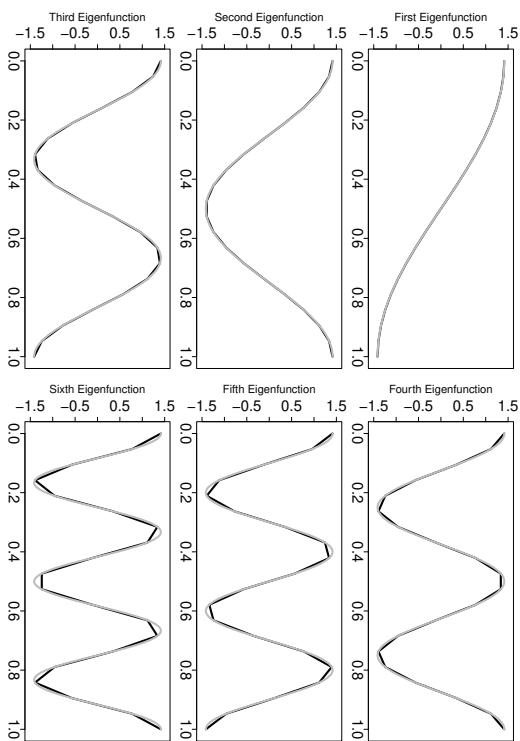


Figure 14: Eigenfunctions of the Cramér-von Mises kernel based on a trapezium rule (TR) with 20 points.

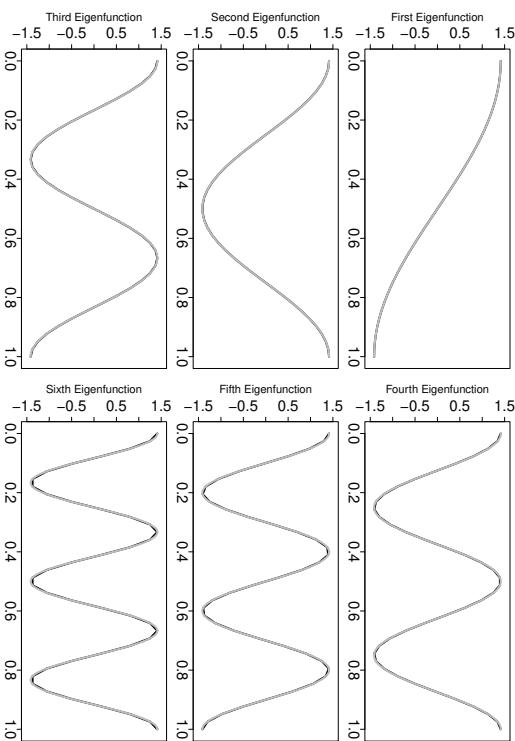


Figure 15: Eigenfunctions of the Cramér-von Mises kernel based on a trapezium rule (TR) with 40 points.

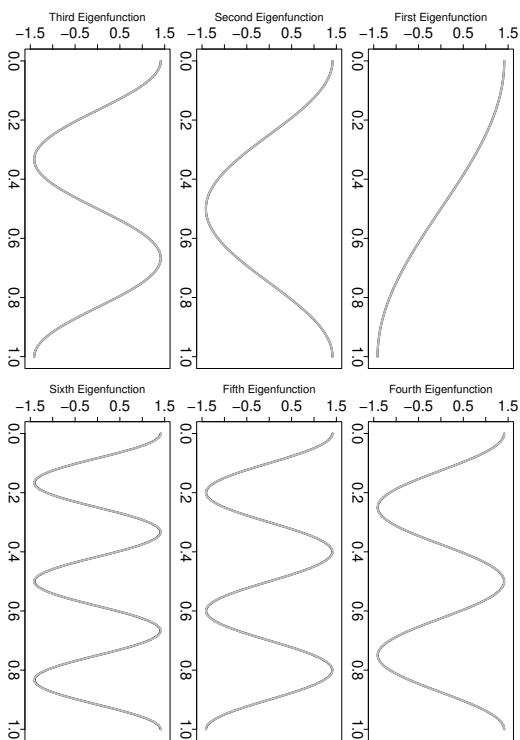


Figure 16: Eigenfunctions of the Cramér-von Mises kernel based on a trapezium rule (TR) with 80 points.

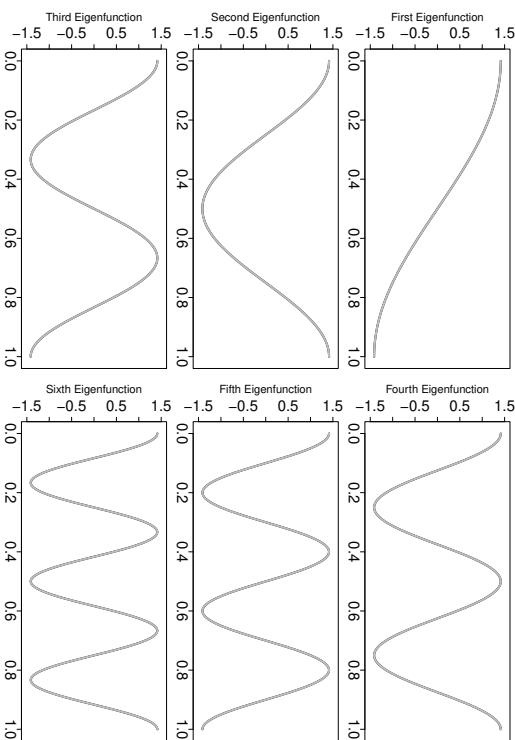


Figure 17: Eigenfunctions of the Cramér-von Mises kernel based on a trapezium rule (TR) with 160 points.

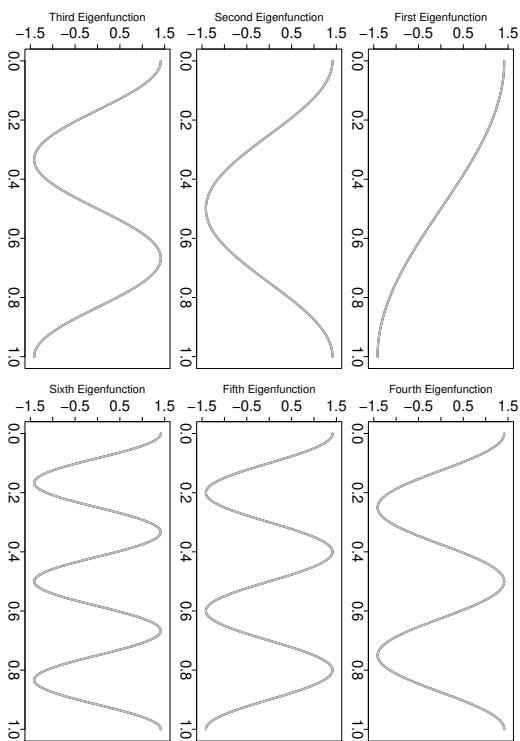


Figure 18: Eigenfunctions of the Cramér-von Mises kernel based on a trapezium rule (TR) with 320 points.

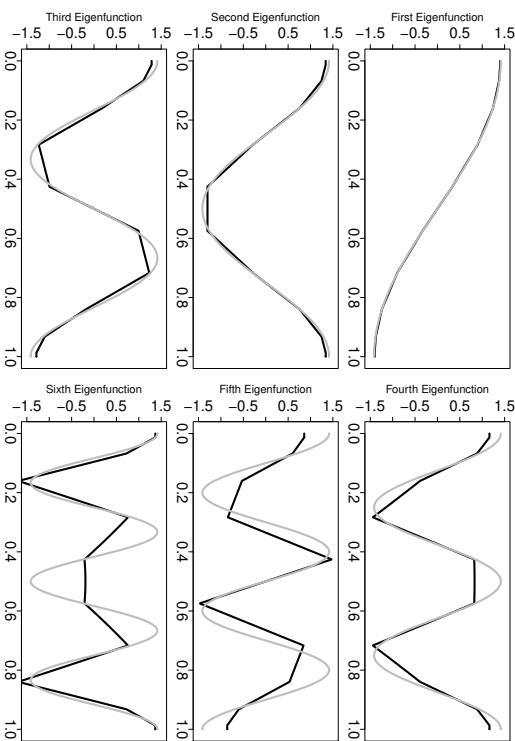


Figure 19: Eigenfunctions of the Cramér-von Mises kernel based on a Gauss-Legendre quadrature rule (GL) with 10 points.

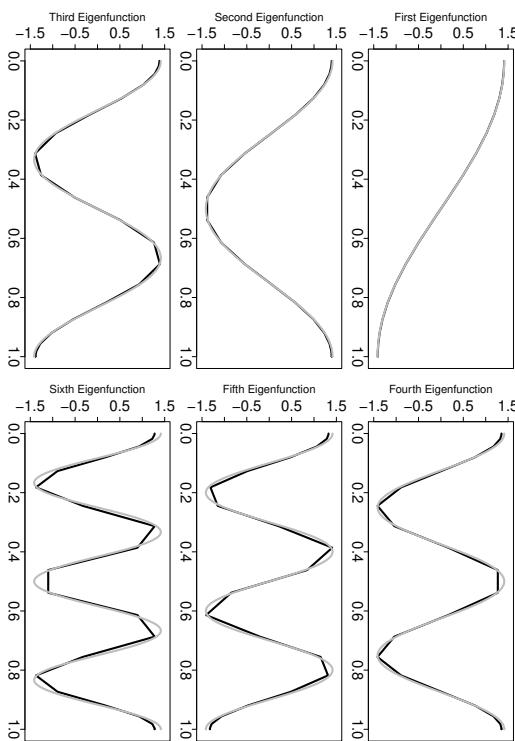


Figure 20: Eigenfunctions of the Cramér-von Mises kernel based on a Gauss-Legendre quadrature rule (GL) with 20 points.

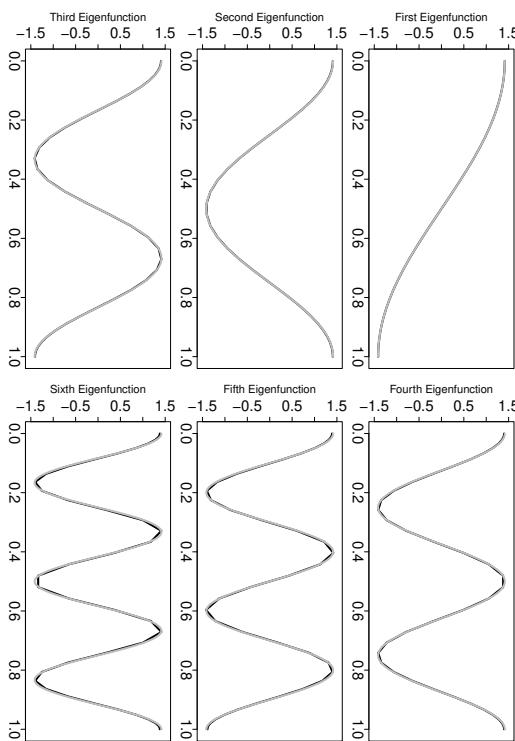


Figure 21: Eigenfunctions of the Cramér-von Mises kernel based on a Gauss-Legendre quadrature rule (GL) with 40 points.

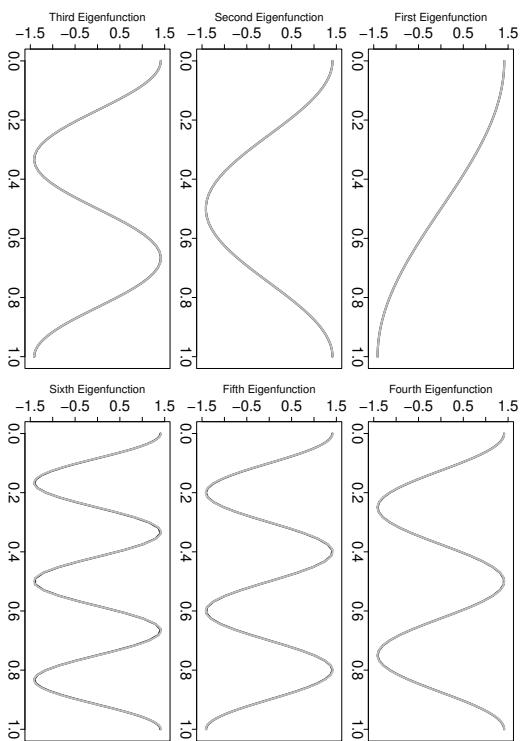


Figure 22: Eigenfunctions of the Cramér-von Mises kernel based on a Gauss-Legendre quadrature rule (GL) with 80 points.

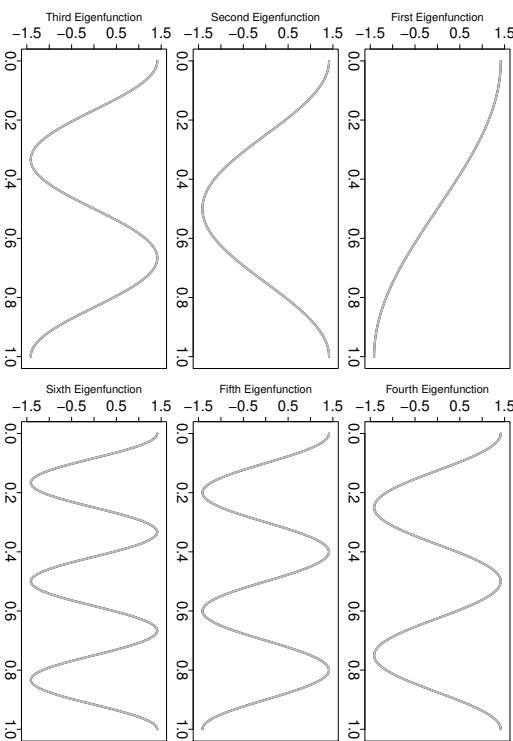


Figure 23: Eigenfunctions of the Cramér-von Mises kernel based on a Gauss-Legendre quadrature rule (GL) with 160 points.

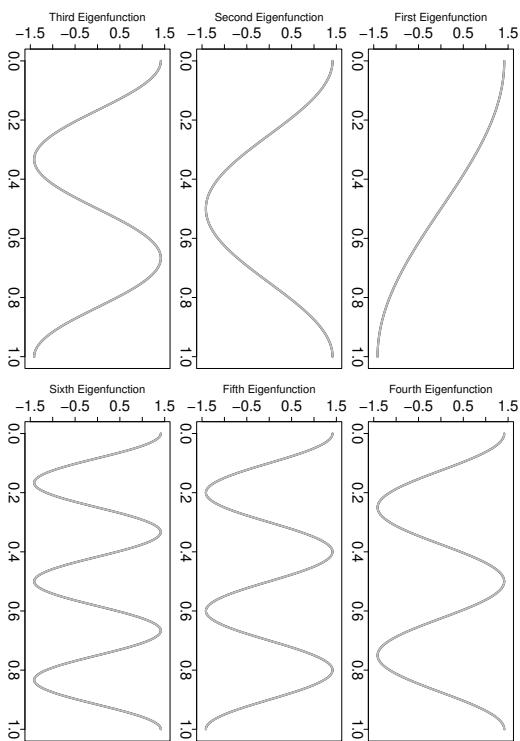


Figure 24: Eigenfunctions of the Cramér-von Mises kernel based on a Gauss-Legendre quadrature rule (GL) with 320 points.

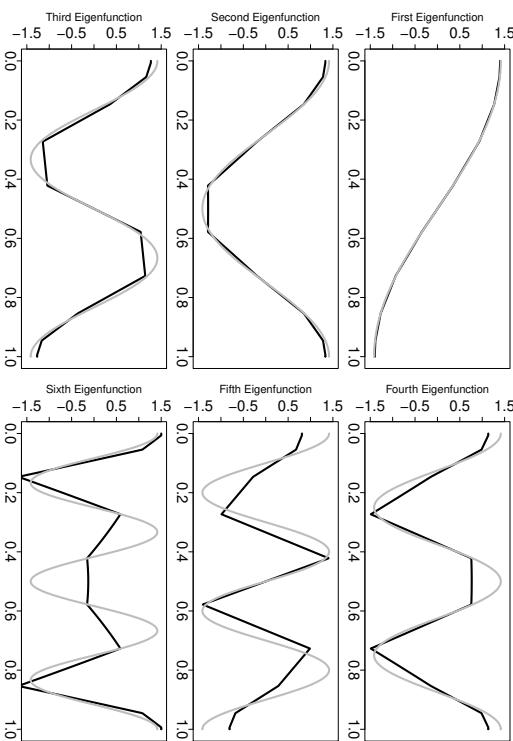


Figure 25: Eigenfunctions of the Cramér-von Mises kernel based on a Clenshaw-Curtis quadrature rule (CC) with 10 points.

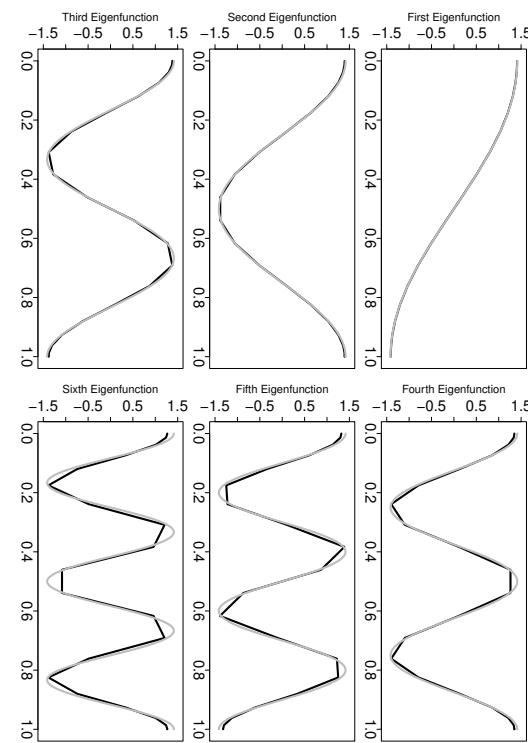


Figure 26: Eigenfunctions of the Cramér-von Mises kernel based on a Clenshaw-Curtis quadrature rule (CC) with 20 points.

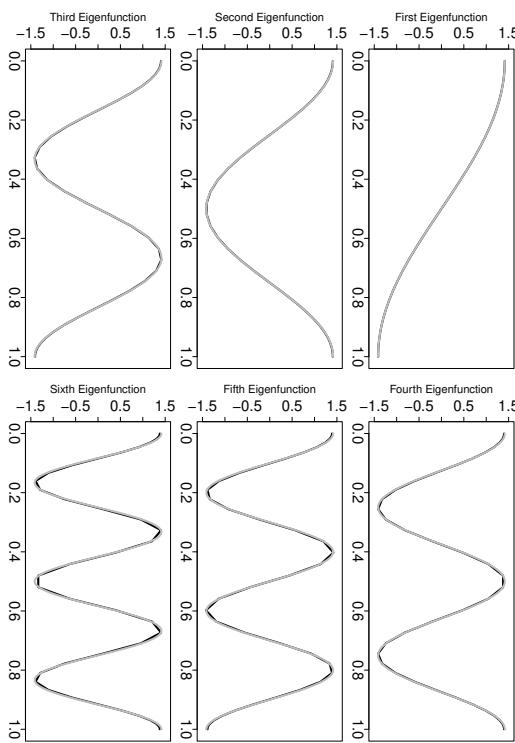


Figure 27: Eigenfunctions of the Cramér-von Mises kernel based on a Clenshaw-Curtis quadrature rule (CC) with 40 points.

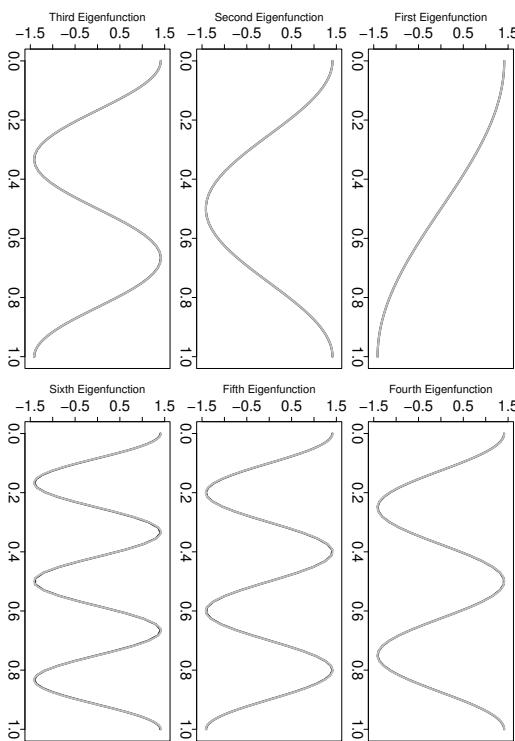


Figure 28: Eigenfunctions of the Cramér-von Mises kernel based on a Clenshaw-Curtis quadrature rule (CC) with 80 points.

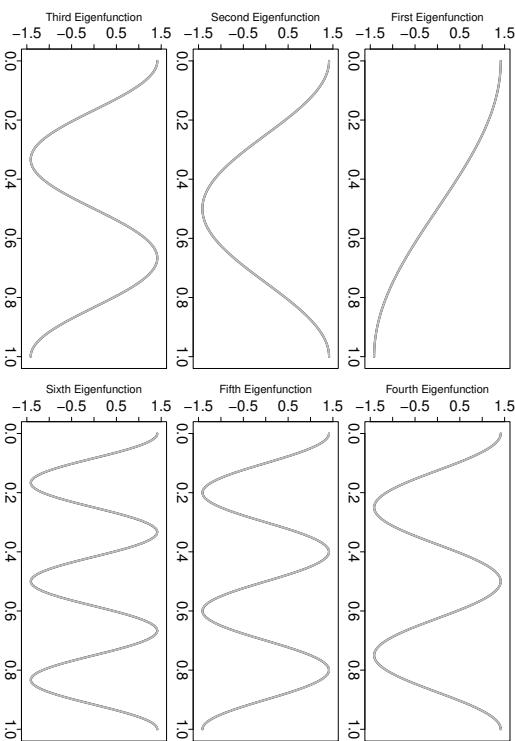


Figure 29: Eigenfunctions of the Cramér-von Mises kernel based on a Clenshaw-Curtis quadrature rule (CC) with 160 points.

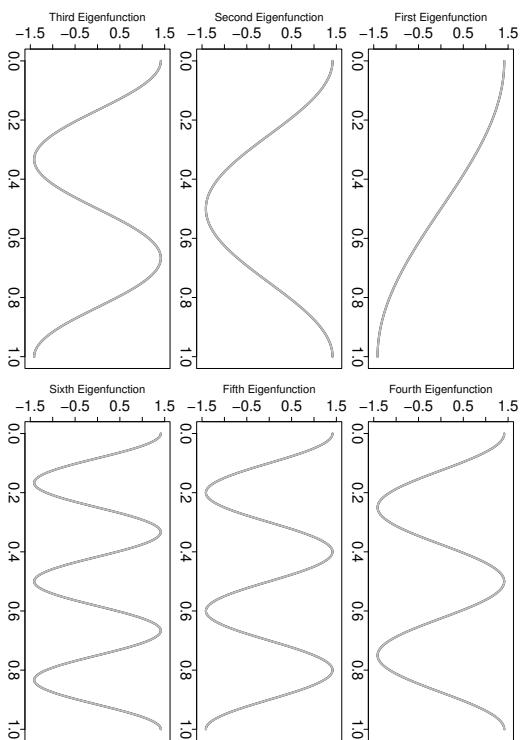


Figure 30: Eigenfunctions of the Gramé-von Mises kernel based on a Clenshaw-Curtis quadrature rule (CC) with 320 points.

1.2. Anderson-Darling Kernel

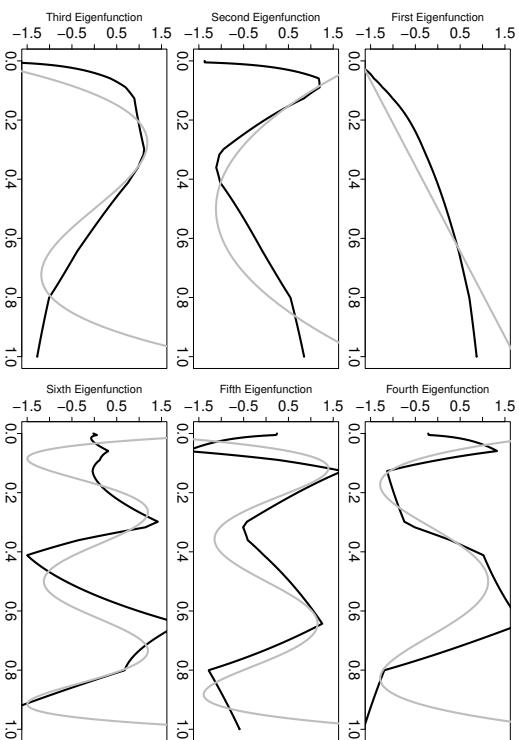


Figure 31: Eigenfunctions of the Anderson-Darling kernel based on a Monte Carlo method (MC) with 10 points.

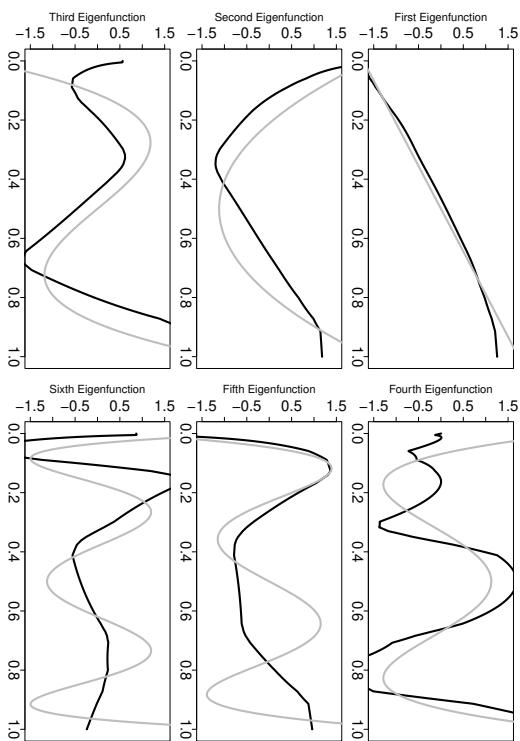


Figure 32: Eigenfunctions of the Anderson-Darling kernel based on a Monte Carlo method (MC) with 20 points.

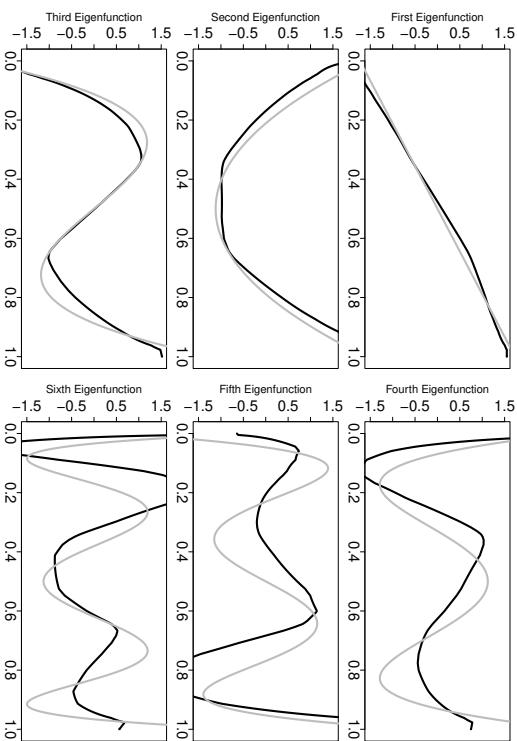


Figure 33: Eigenfunctions of the Anderson-Darling kernel based on a Monte Carlo method (MC) with 40 points.

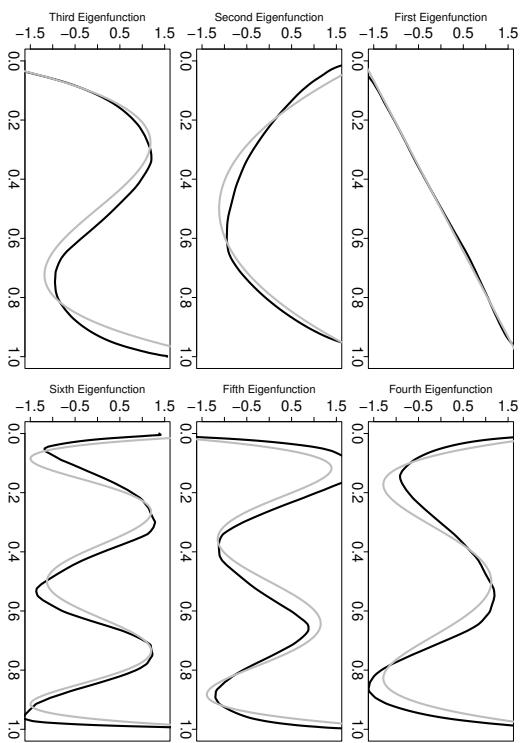


Figure 34: Eigenfunctions of the Anderson-Darling kernel based on a Monte Carlo method (MC) with 80 points.

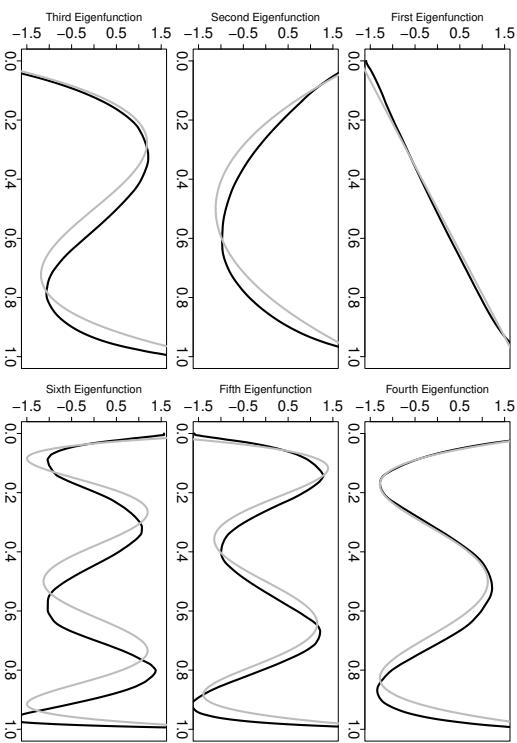


Figure 35: Eigenfunctions of the Anderson-Darling kernel based on a Monte Carlo method (MC) with 160 points.

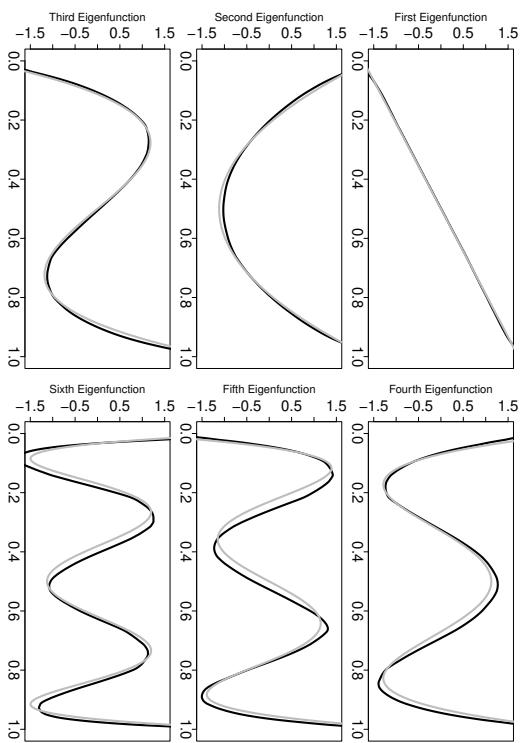


Figure 36: Eigenfunctions of the Anderson-Darling kernel based on a Monte Carlo method (MC) with 320 points.

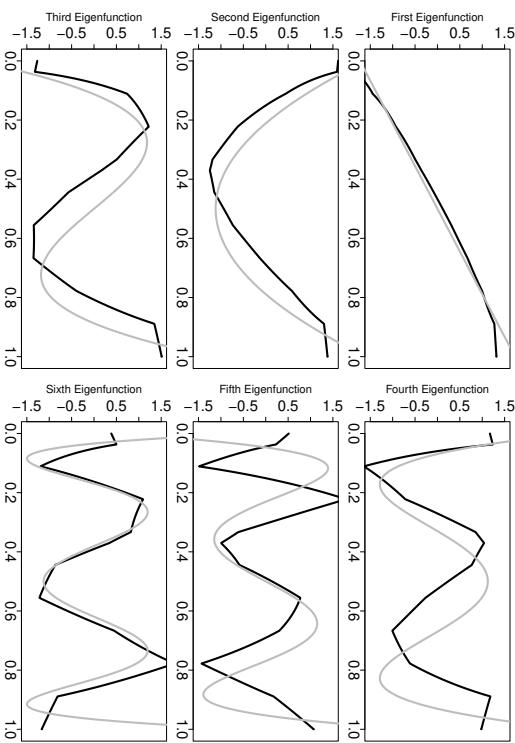


Figure 37: Eigenfunctions of the Anderson-Darling kernel based on a quasi-Monte Carlo method (HA) with 10 points.

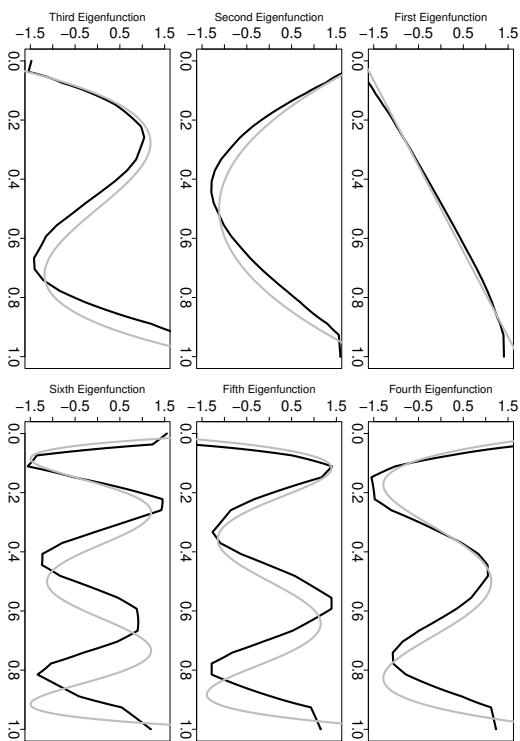


Figure 38: Eigenfunctions of the Anderson-Darling kernel based on a quasi-Monte Carlo method (HA) with 20 points.

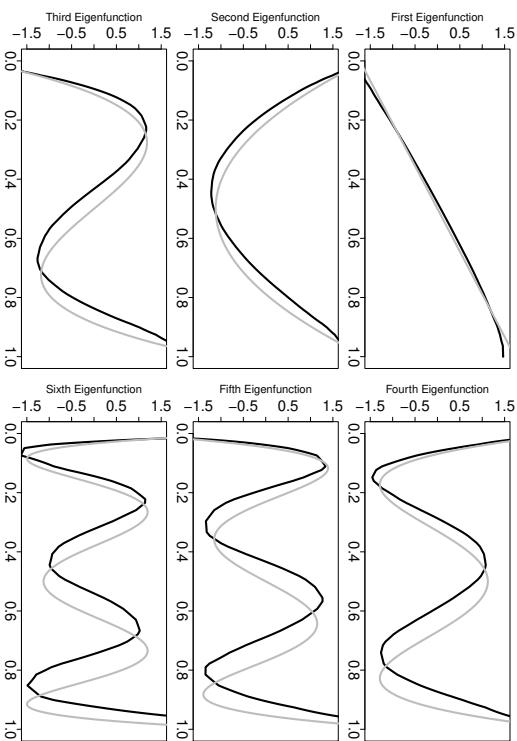


Figure 39: Eigenfunctions of the Anderson-Darling kernel based on a quasi-Monte Carlo method (HA) with 40 points.

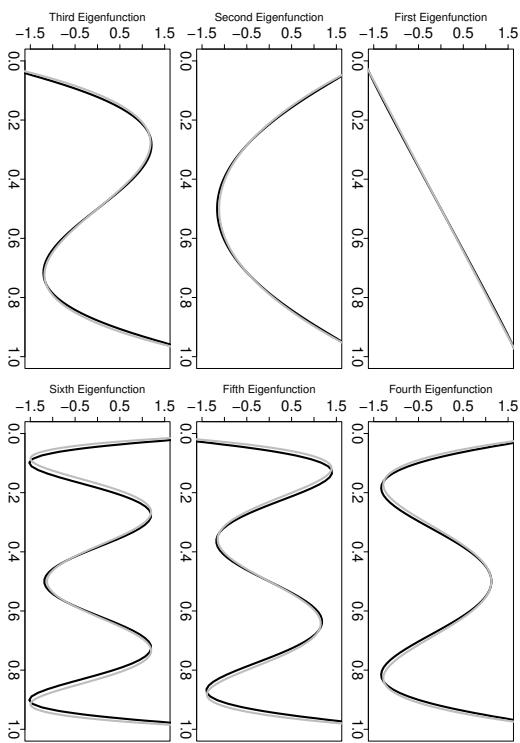


Figure 40: Eigenfunctions of the Anderson-Darling kernel based on a quasi-Monte Carlo method (HA) with 80 points.

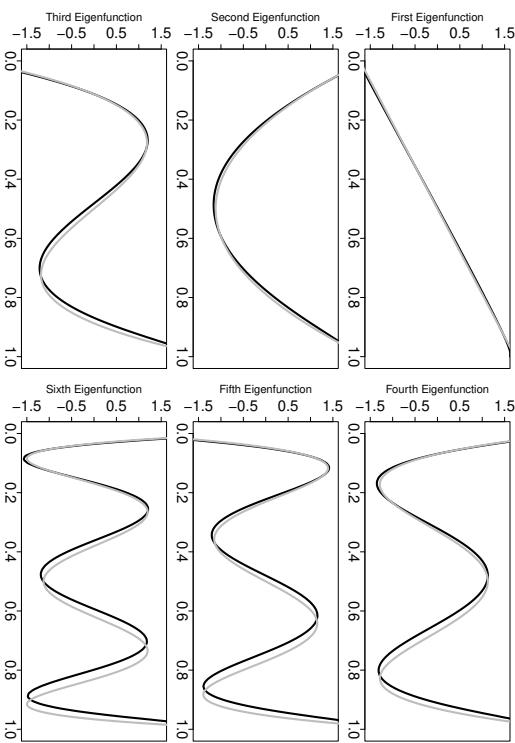


Figure 41: Eigenfunctions of the Anderson-Darling kernel based on a quasi-Monte Carlo method (HA) with 160 points.

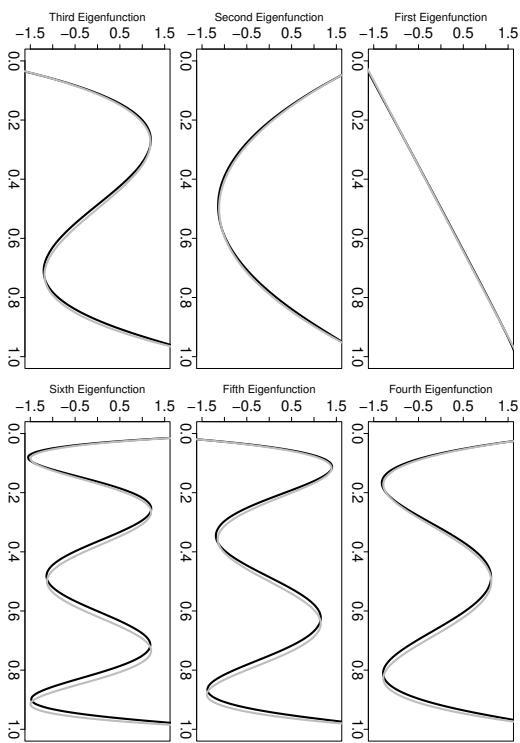


Figure 42: Eigenfunctions of the Anderson-Darling kernel based on a quasi-Monte Carlo method (HA) with 320 points.

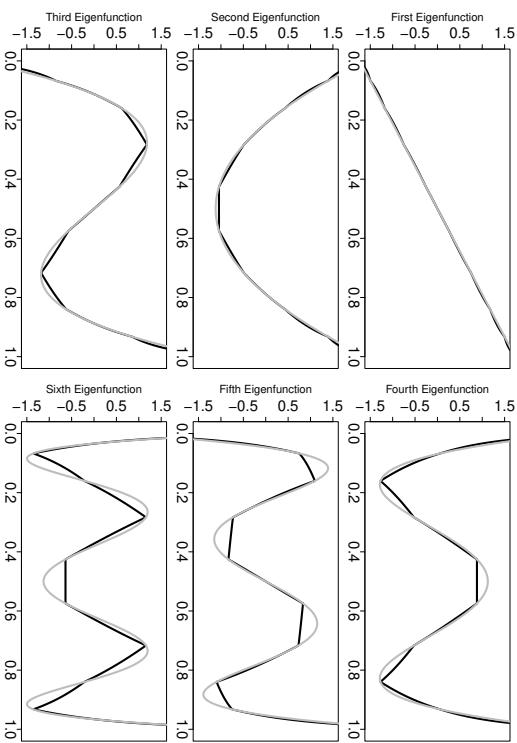


Figure 43: Eigenfunctions of the Anderson-Darling kernel based on a Gauss-Legendre quadrature rule (GL) with 10 points.

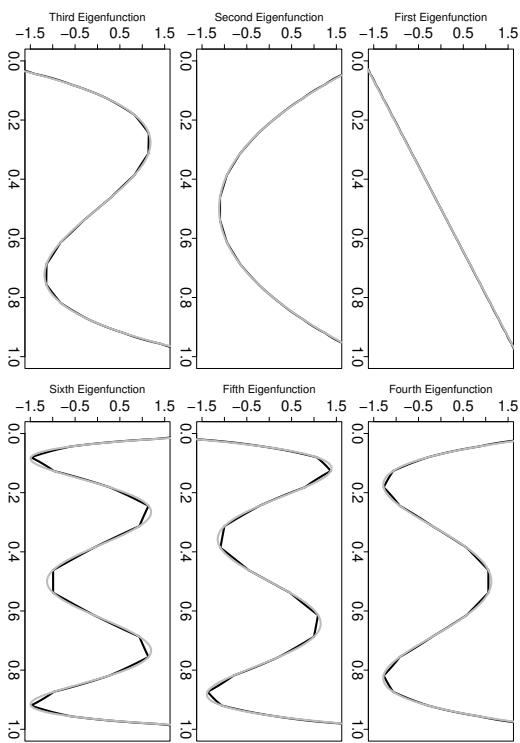


Figure 44: Eigenfunctions of the Anderson-Darling kernel based on a Gauss-Legendre quadrature rule (GL) with 20 points.

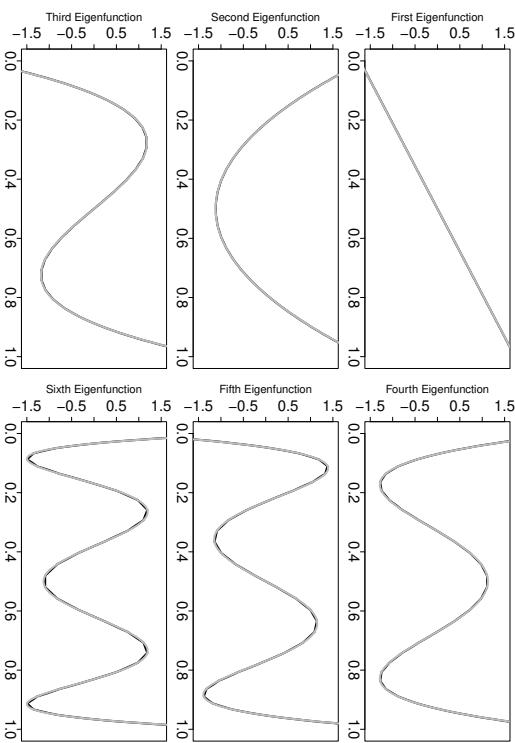


Figure 45: Eigenfunctions of the Anderson-Darling kernel based on a Gauss-Legendre quadrature rule (GL) with 40 points.

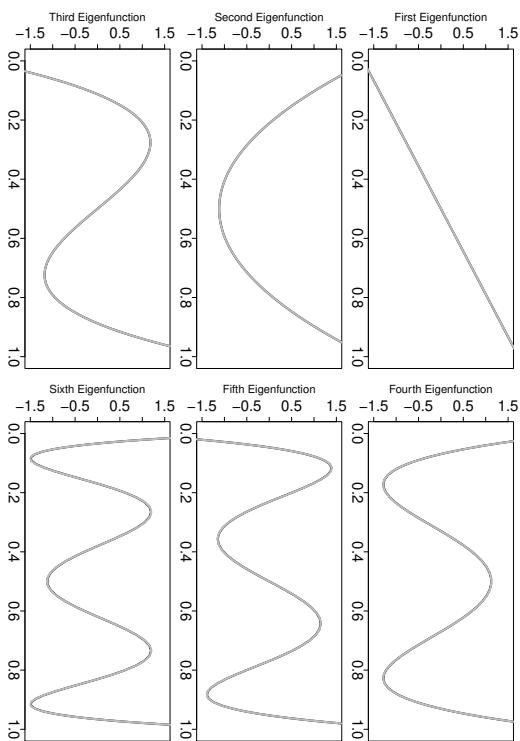


Figure 46: Eigenfunctions of the Anderson-Darling kernel based on a Gauss-Legendre quadrature rule (GL) with 80 points.

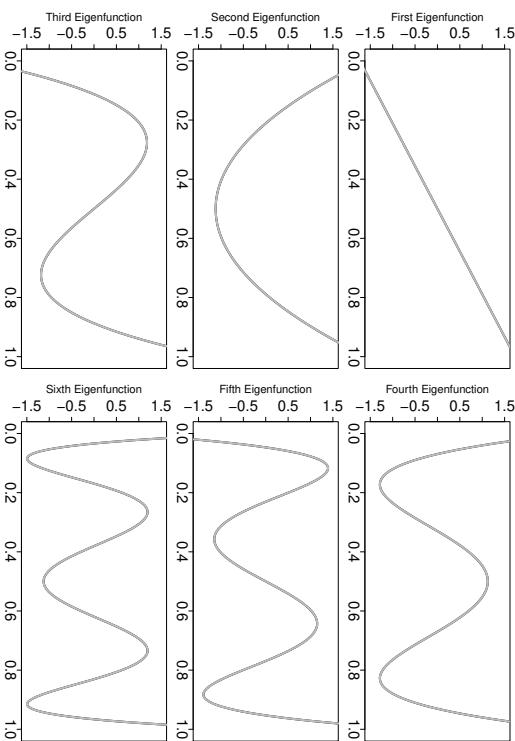


Figure 47: Eigenfunctions of the Anderson-Darling kernel based on a Gauss-Legendre quadrature rule (GL) with 160 points.

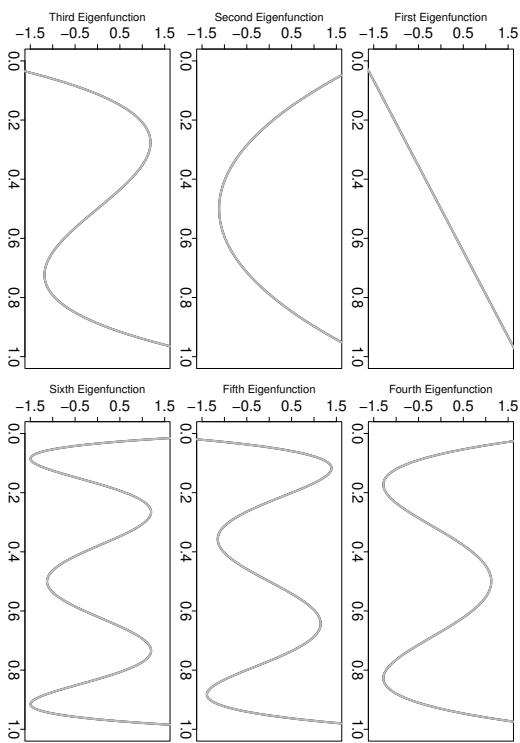


Figure 48: Eigenfunctions of the Anderson-Darling kernel based on a Gauss-Legendre quadrature rule (GL) with 320 points.

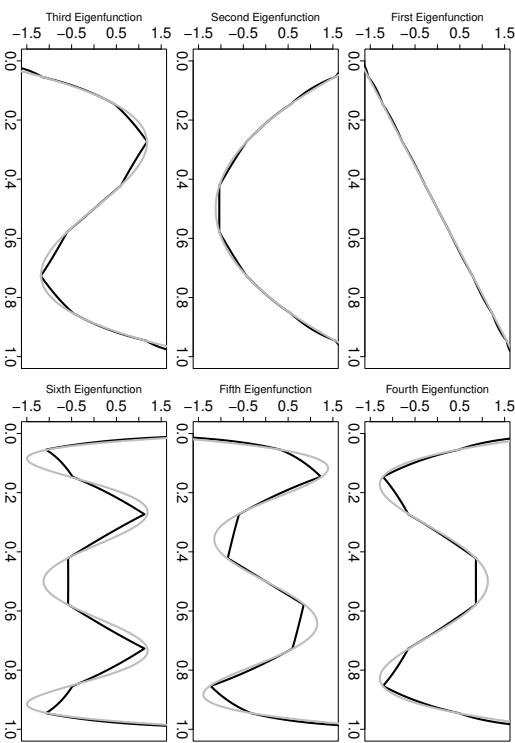


Figure 49: Eigenfunctions of the Anderson-Darling kernel based on a Clenshaw-Curtis quadrature rule (CC) with 10 points.

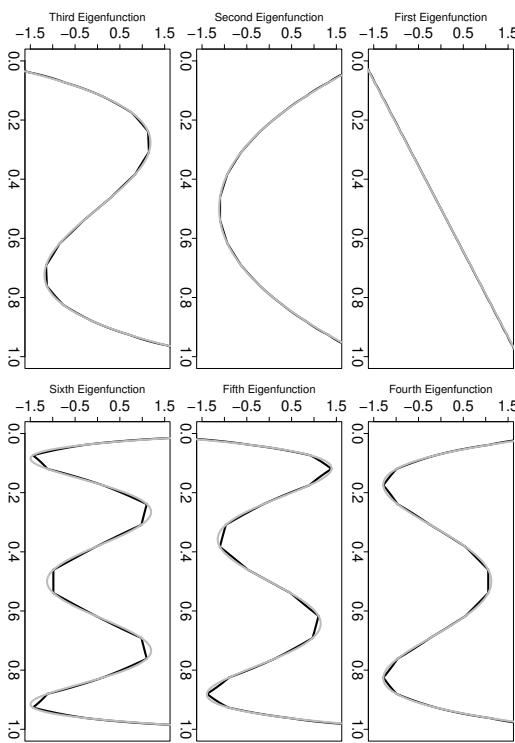


Figure 50: Eigenfunctions of the Anderson-Darling kernel based on a Clenshaw-Curtis quadrature rule (CC) with 20 points.

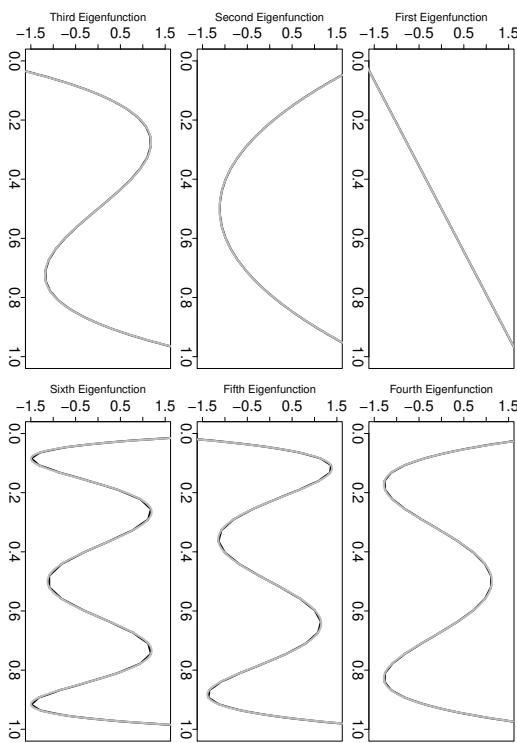


Figure 51: Eigenfunctions of the Anderson-Darling kernel based on a Clenshaw-Curtis quadrature rule (CC) with 40 points.

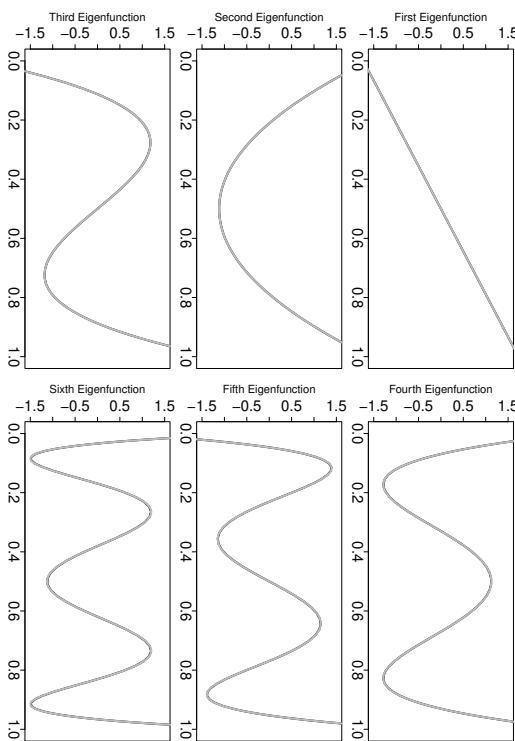


Figure 52: Eigenfunctions of the Anderson-Darling kernel based on a Clenshaw-Curtis quadrature rule (CC) with 80 points.

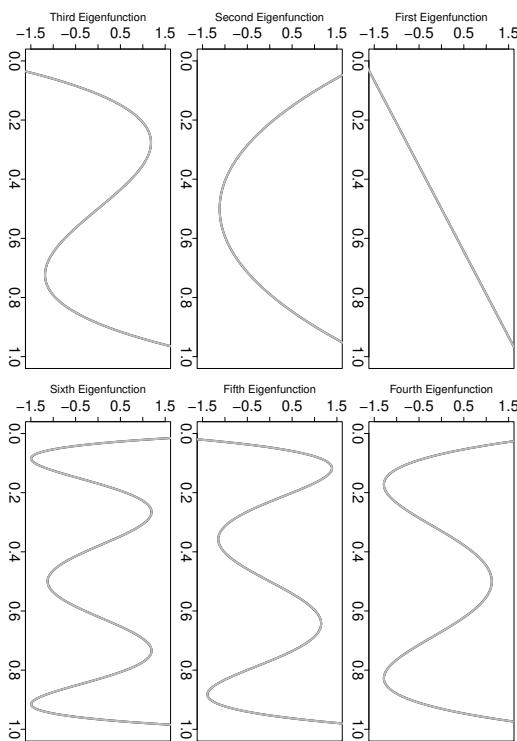


Figure 53: Eigenfunctions of the Anderson-Darling kernel based on a Clenshaw-Curtis quadrature rule (CC) with 160 points.

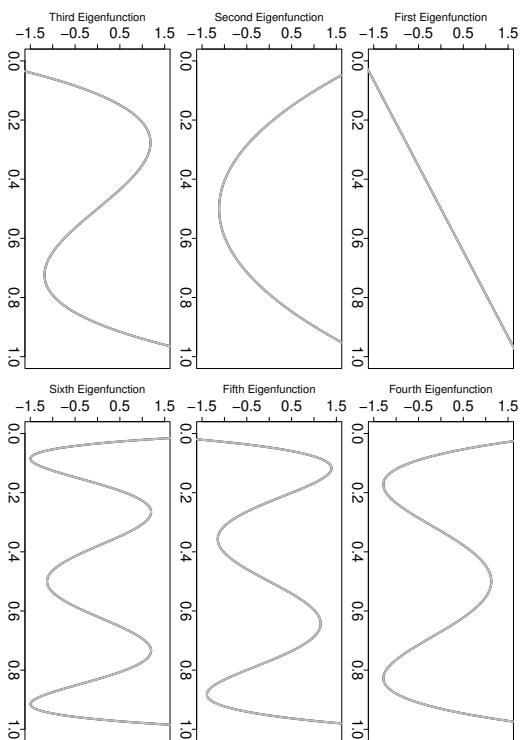


Figure 54: Eigenfunctions of the Anderson-Darling kernel based on a Clenshaw-Curtis quadrature rule (CC) with 320 points.

1.3. Watson Kernel

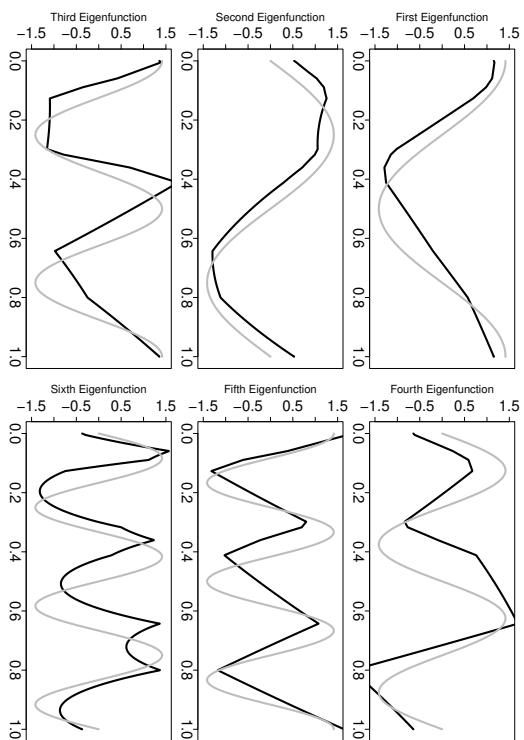


Figure 55: Eigenfunctions of the Watson kernel based on a Monte Carlo method (MC) with 10 points.

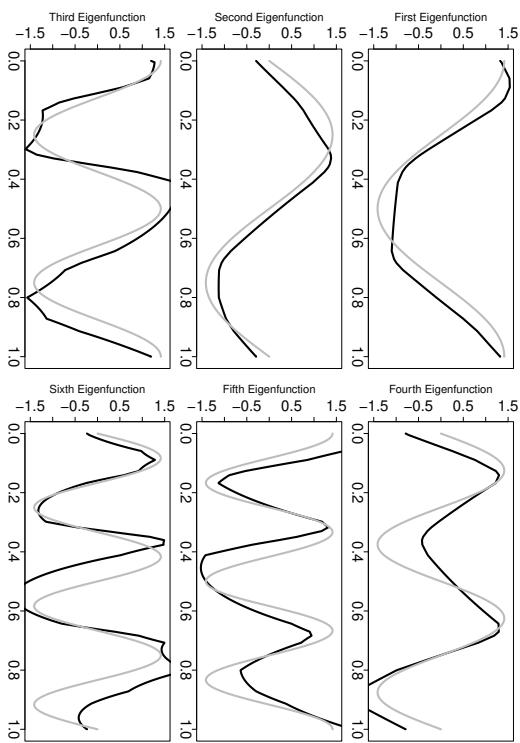


Figure 56: Eigenfunctions of the Watson kernel based on a Monte Carlo method (MC) with 20 points.

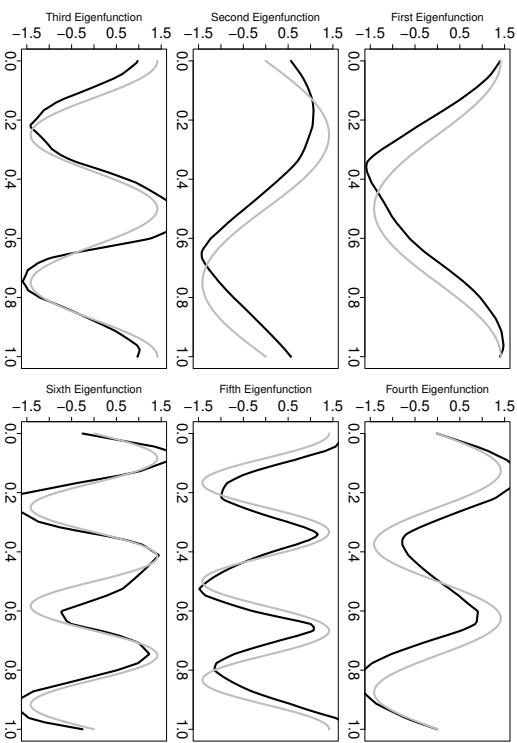


Figure 57: Eigenfunctions of the Watson kernel based on a Monte Carlo method (MC) with 40 points.

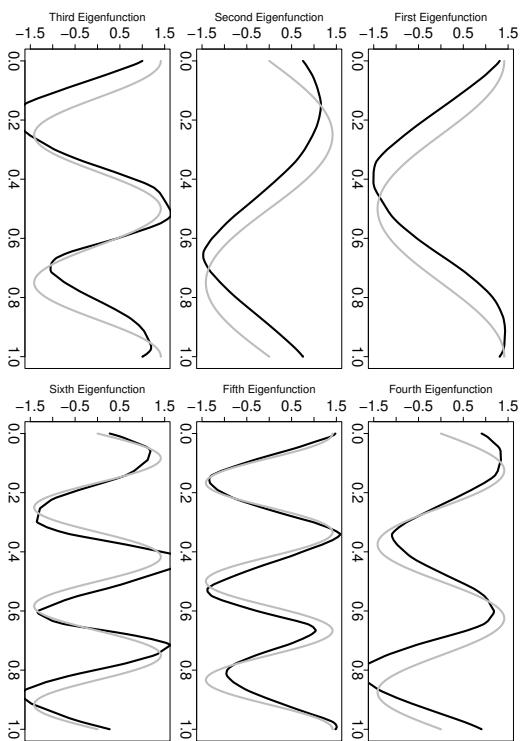


Figure 58: Eigenfunctions of the Watson kernel based on a Monte Carlo method (MC) with 80 points.

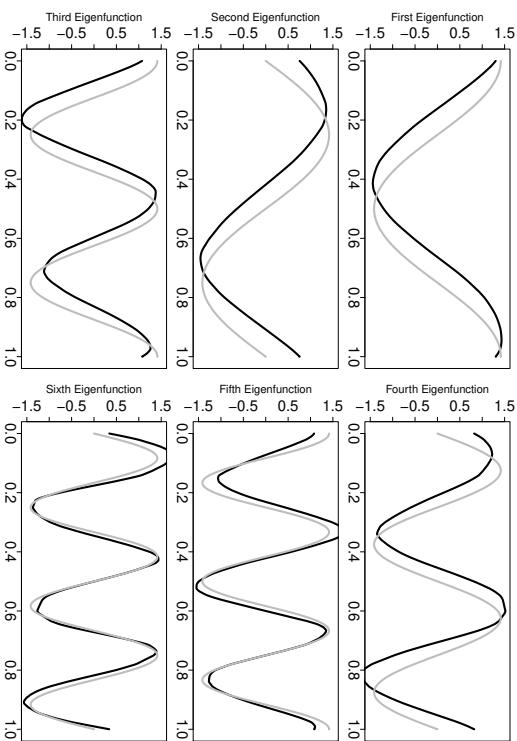


Figure 59: Eigenfunctions of the Watson kernel based on a Monte Carlo method (MC) with 160 points.

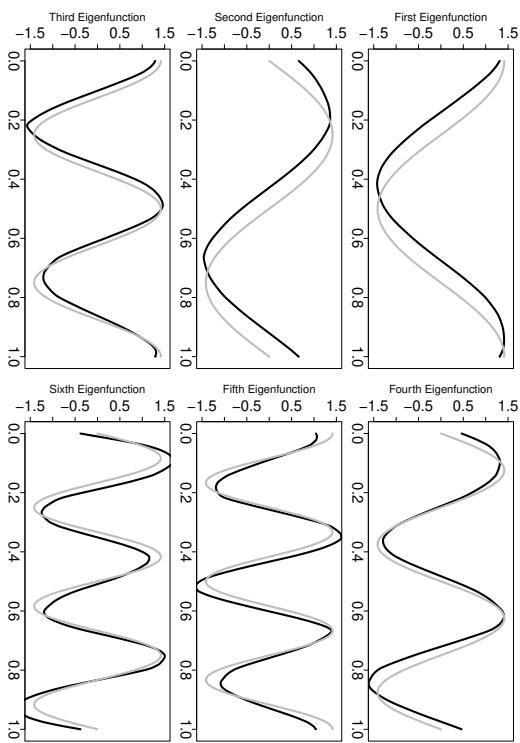


Figure 60: Eigenfunctions of the Watson kernel based on a Monte Carlo method (MC) with 320 points.

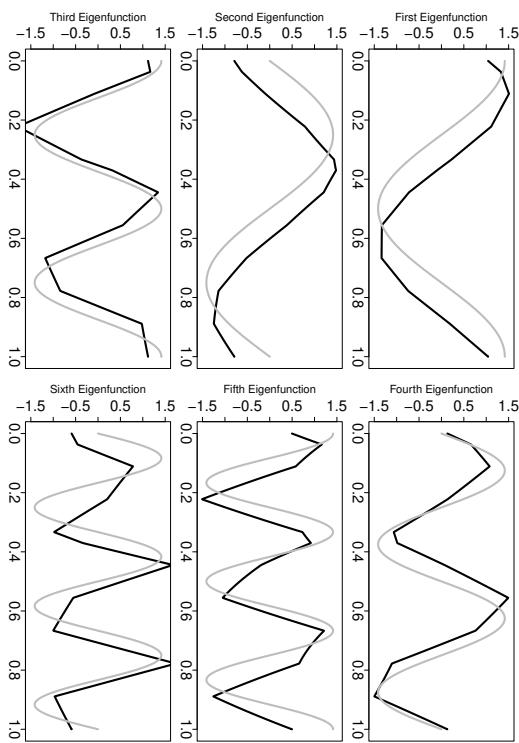


Figure 61: Eigenfunctions of the Watson kernel based on a quasi-Monte Carlo method (HA) with 10 points.

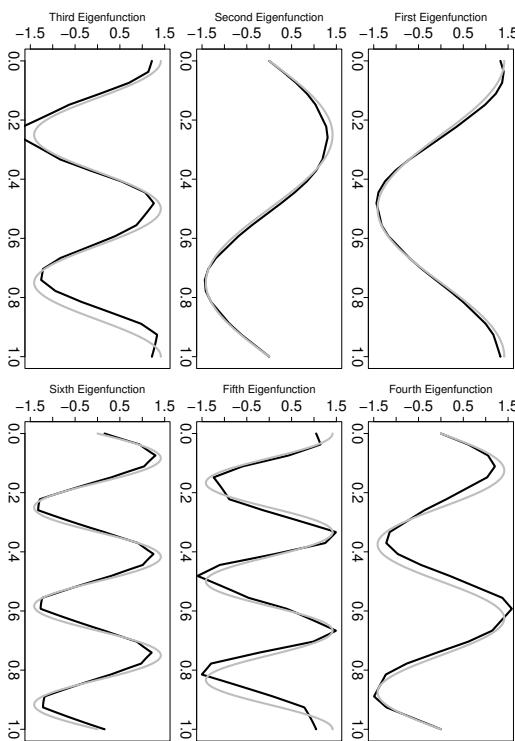


Figure 62: Eigenfunctions of the Watson kernel based on a quasi-Monte Carlo method (HA) with 20 points.

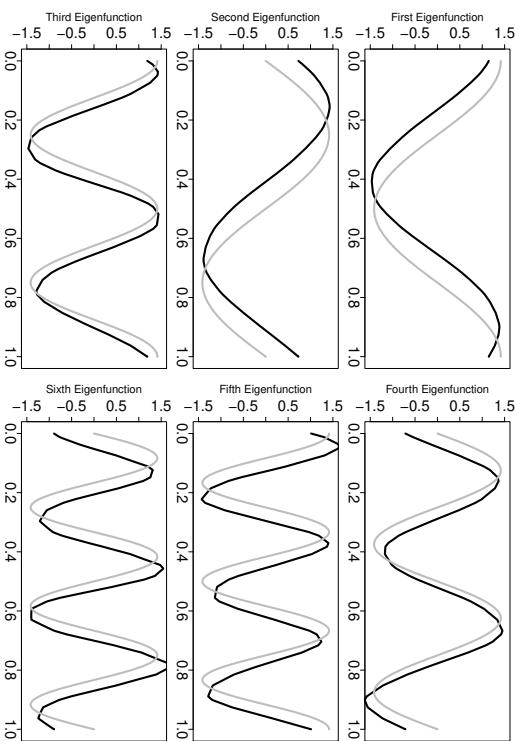


Figure 63: Eigenfunctions of the Watson kernel based on a quasi-Monte Carlo method (HA) with 40 points.

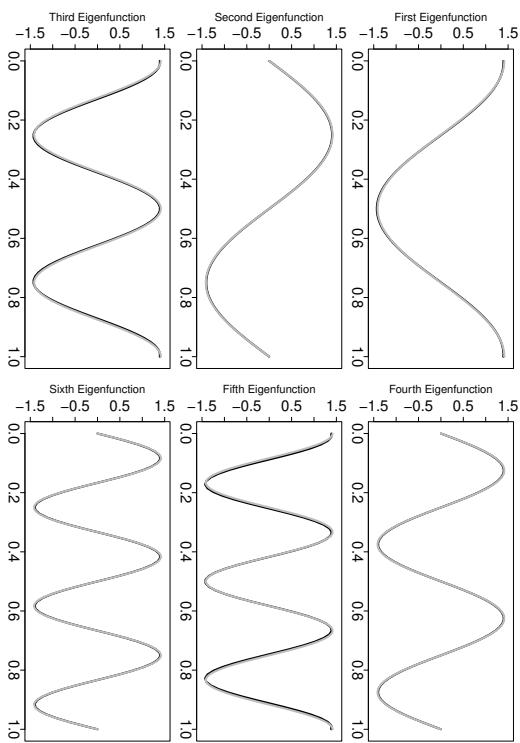


Figure 64: Eigenfunctions of the Watson kernel based on a quasi-Monte Carlo method (HA) with 80 points.

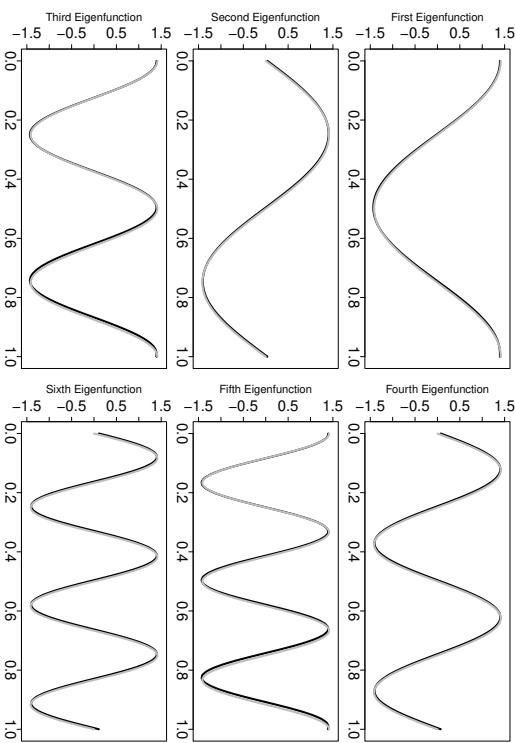


Figure 65: Eigenfunctions of the Watson kernel based on a quasi-Monte Carlo method (HA) with 160 points.

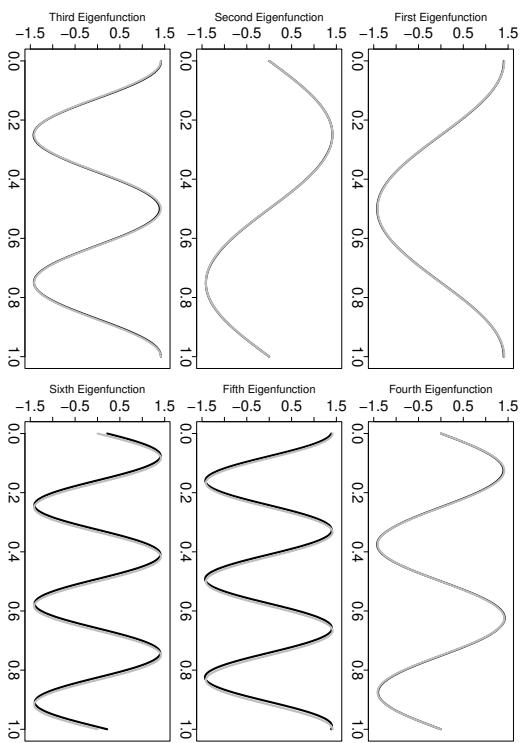


Figure 66: Eigenfunctions of the Watson kernel based on a quasi-Monte Carlo method (HA) with 320 points.

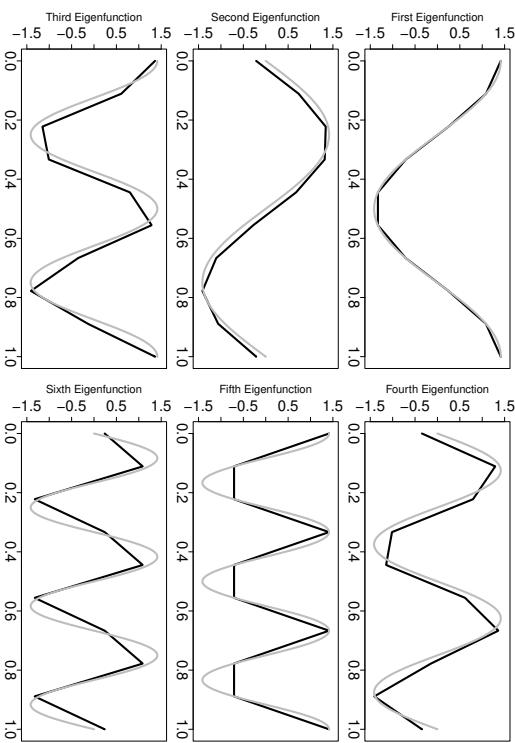


Figure 67: Eigenfunctions of the Watson kernel based on a trapezium rule (TR) with 10 points.

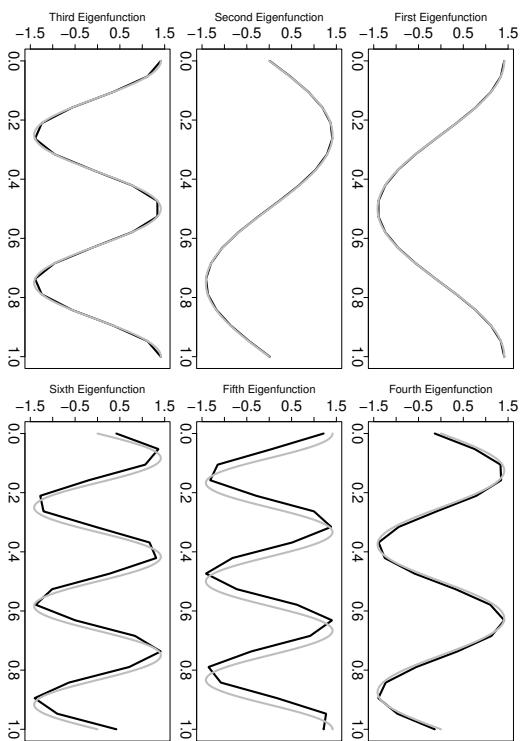


Figure 68: Eigenfunctions of the Watson kernel based on a trapezium rule (TR) with 20 points.

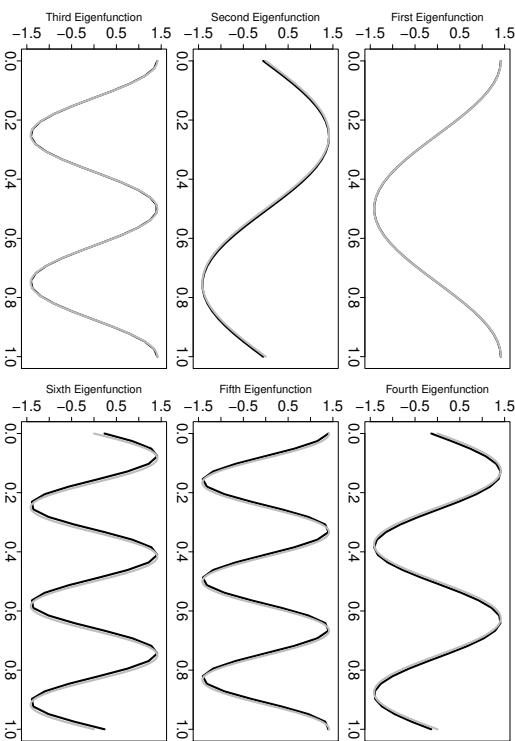


Figure 69: Eigenfunctions of the Watson kernel based on a trapezium rule (TR) with 40 points.

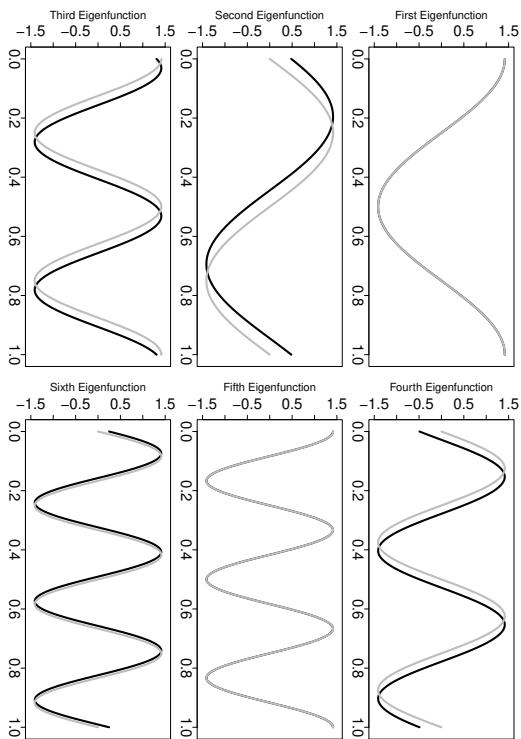


Figure 71: Eigenfunctions of the Watson kernel based on a trapezium rule (TR) with 160 points.

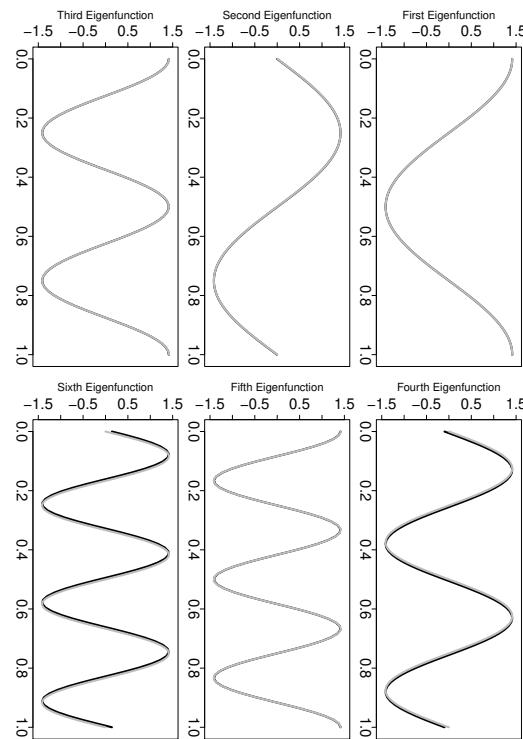


Figure 70: Eigenfunctions of the Watson kernel based on a trapezium rule (TR) with 80 points.

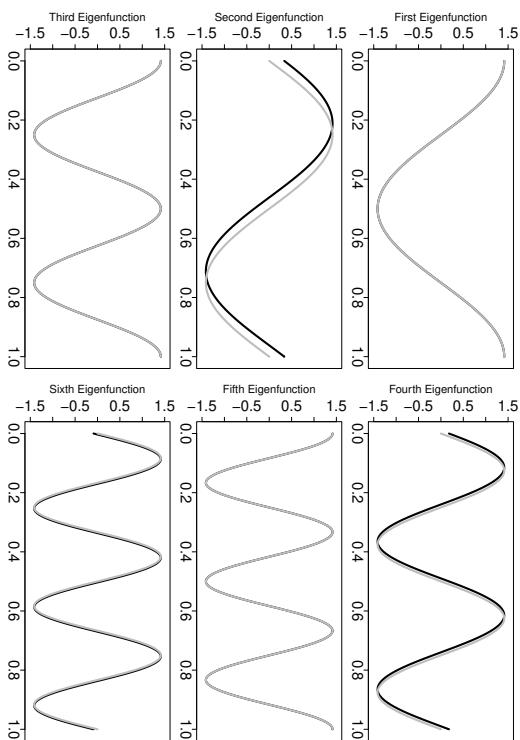


Figure 72: Eigenfunctions of the Watson kernel based on a trapezium rule (TR) with 320 points.

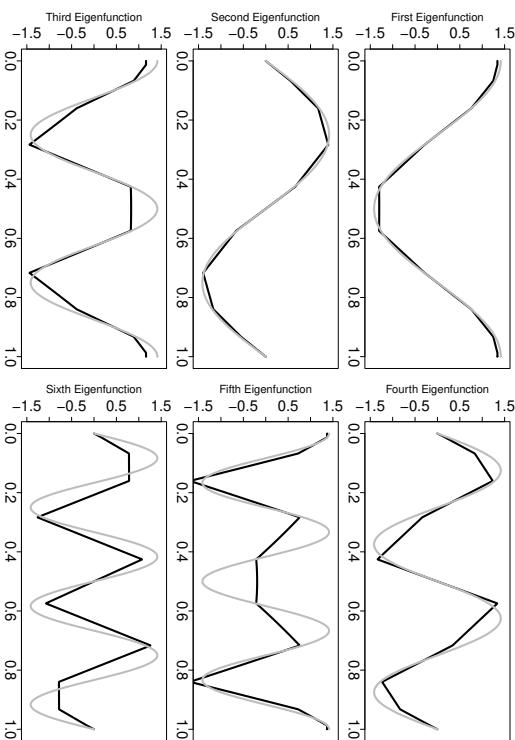


Figure 73: Eigenfunctions of the Watson kernel based on a Gauss-Legendre quadrature rule (GL) with 10 points.

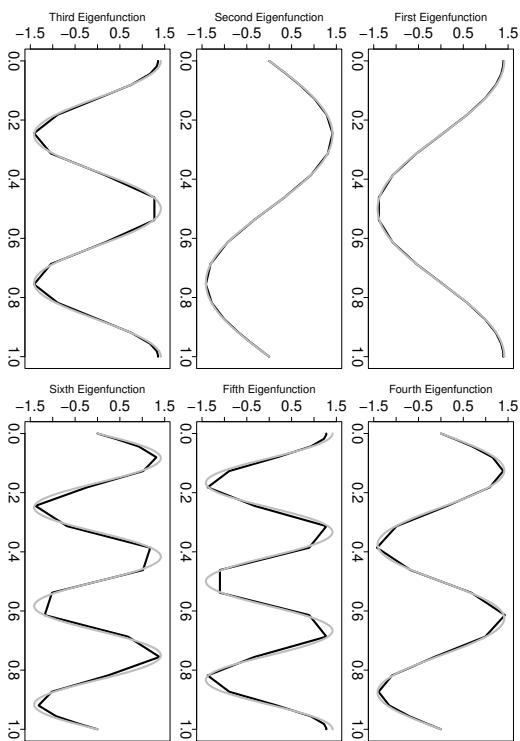


Figure 74: Eigenfunctions of the Watson kernel based on a Gauss-Legendre quadrature rule (GL) with 20 points.

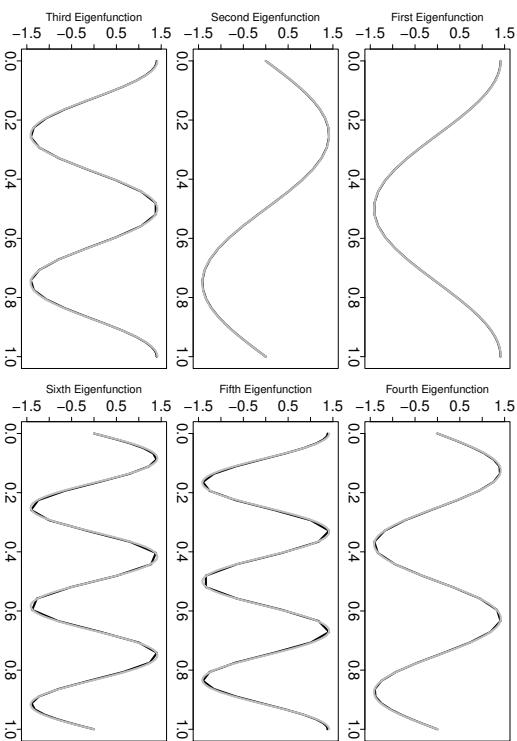


Figure 75: Eigenfunctions of the Watson kernel based on a Gauss-Legendre quadrature rule (GL) with 40 points.

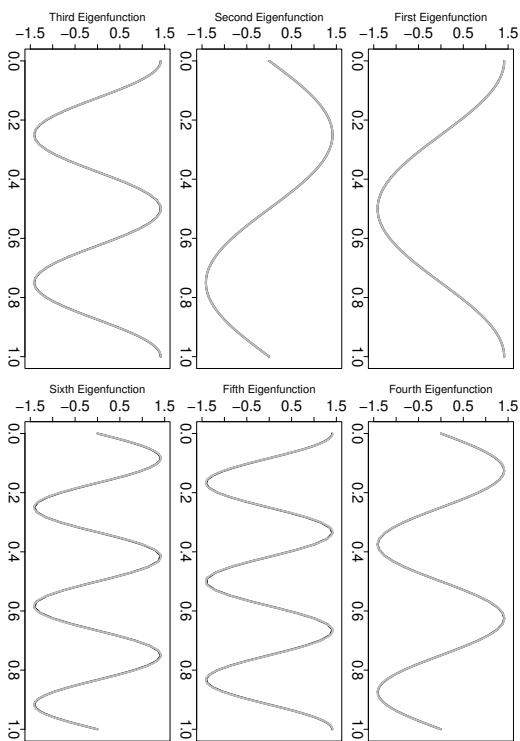


Figure 76: Eigenfunctions of the Watson kernel based on a Gauss-Legendre quadrature rule (GL) with 80 points.

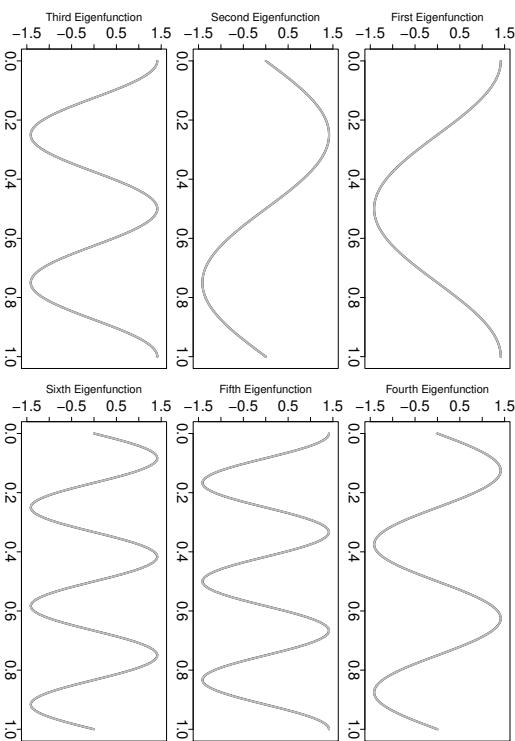


Figure 77: Eigenfunctions of the Watson kernel based on a Gauss-Legendre quadrature rule (GL) with 160 points.

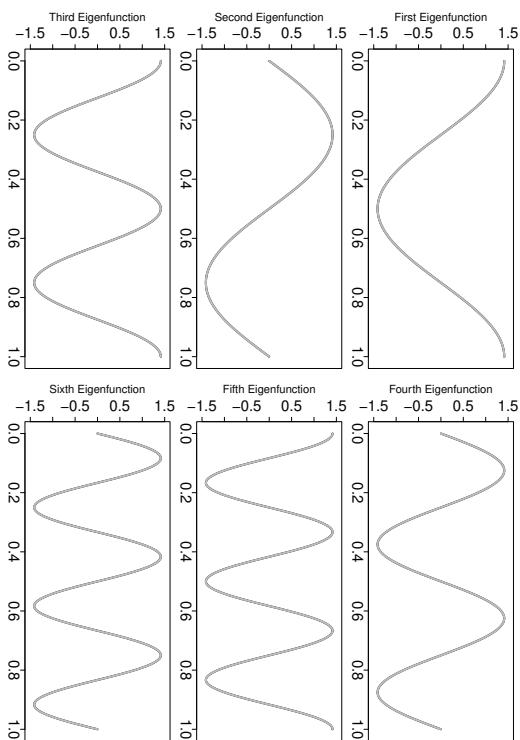


Figure 78: Eigenfunctions of the Watson kernel based on a Gauss-Legendre quadrature rule (GL) with 320 points.

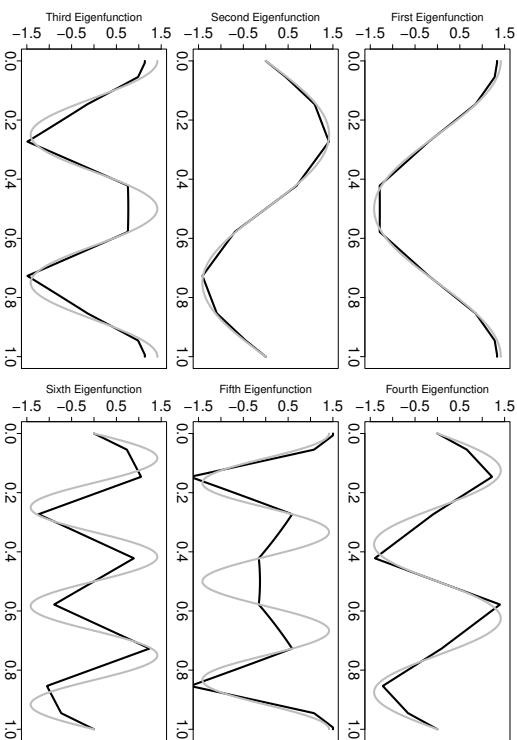


Figure 79: Eigenfunctions of the Watson kernel based on a Clenshaw-Curtis quadrature rule (CC) with 10 points.

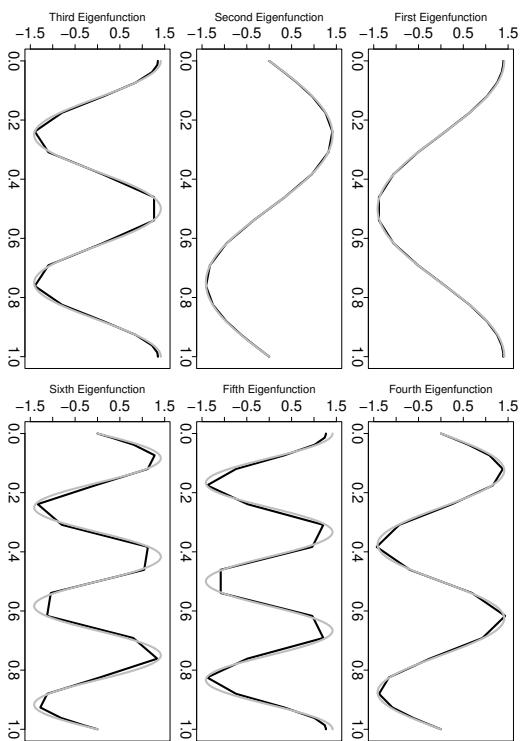


Figure 80: Eigenfunctions of the Watson kernel based on a Clenshaw-Curtis quadrature rule (CC) with 20 points.

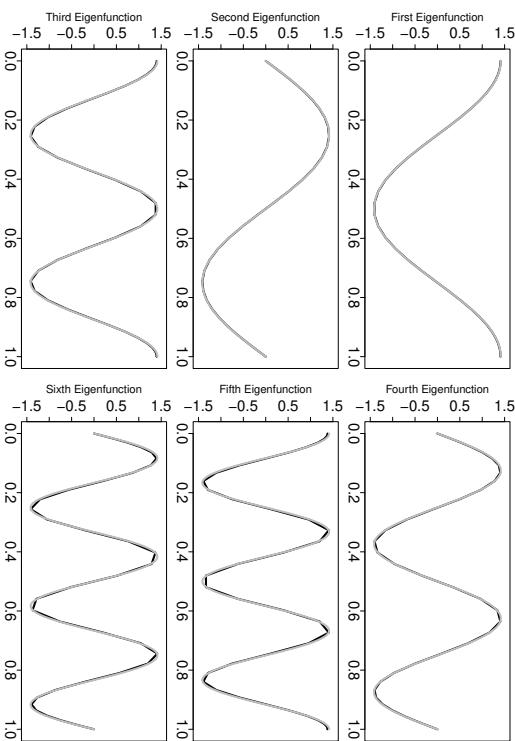


Figure 81: Eigenfunctions of the Watson kernel based on a Clenshaw-Curtis quadrature rule (CC) with 40 points.

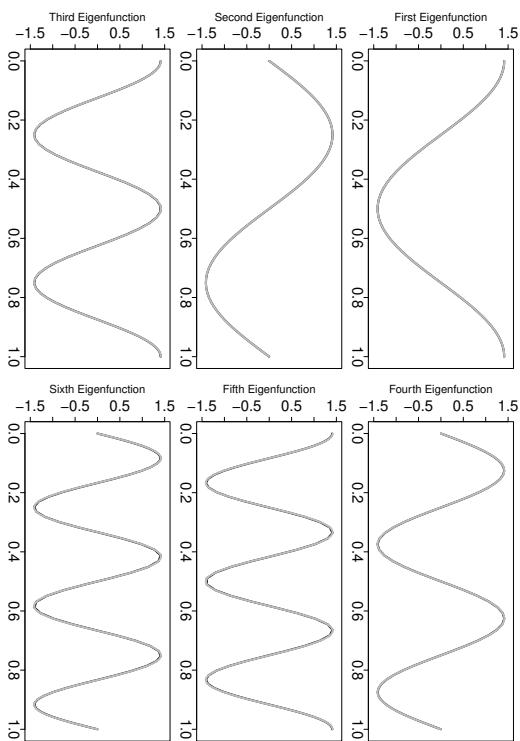


Figure 82: Eigenfunctions of the Watson kernel based on a Clenshaw-Curtis quadrature rule (CC) with 80 points.

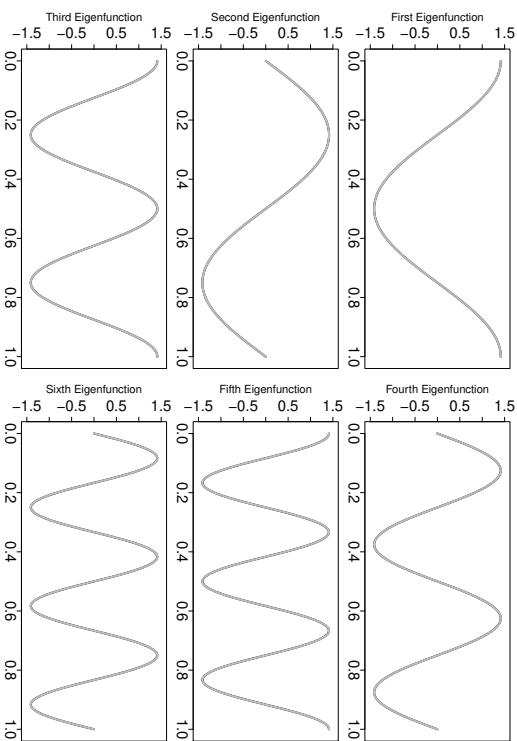


Figure 83: Eigenfunctions of the Watson kernel based on a Clenshaw-Curtis quadrature rule (CC) with 160 points.

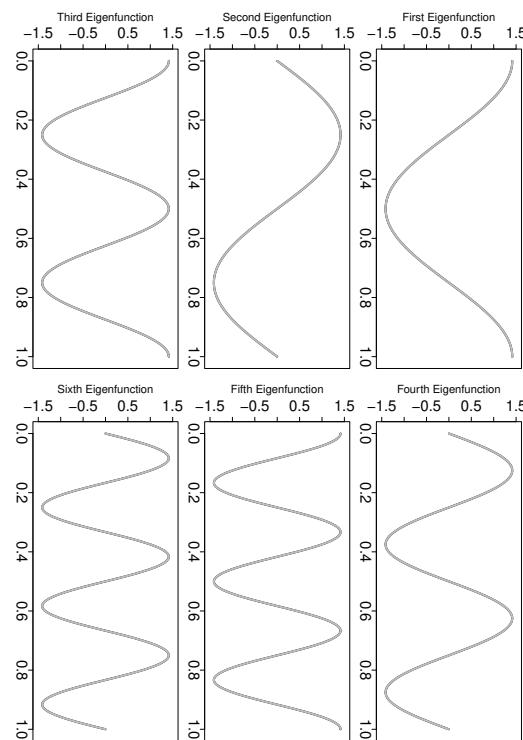


Figure 84: Eigenfunctions of the Watson kernel based on a Clenshaw-Curtis quadrature rule (CC) with 320 points.