Available at WWW.MATHEMATICSWEB.ORG
POWERED BY SCIENCE @ DIRECT•

Journal of Multivariate Analysis 87 (2003) 370–397

Journal of Multivariate Analysis

http://www.elsevier.com/locate/jmva

Tail behaviour of Gaussian processes with applications to the Brownian pillow

Alex J. Koning* and Vladimir Protasov

Econometric Institute, Erasmus University Rotterdam, P.O. Box 1738 NL-3000 DR, Rotterdam, Netherlands

Received 7 January 2002

Abstract

In this paper we investigate the tail behaviour of a random variable S which may be viewed as a functional T of a zero mean Gaussian process X, taking special interest in the situation where X obeys the structure which is typical for limiting processes occurring in nonparametric testing of (multivariate) independency and (multivariate) constancy over time. The tail behaviour of S is described by means of a constant a and a random variable R which is defined on the same probability space as S. The constant a acts as an upper bound, and is relevant for the computation of the efficiency of test statistics converging in distribution to S. The random variable R acts as a lower bound, and is instrumental in deriving approximation for the upper percentage points of S by simulation.

AMS 2000 subject classifications: primary 60G15; secondary 62E20; 62G10

Keywords: Tail behaviour; Gaussian processes; Brownian pillow; Asymptotic distribution theory; Kolmogorov-type tests; Cramér–von Mises type tests; Anderson–Darling-type tests; Multivariate constancy; Multivariate independence

1. Introduction

Let d be an integer greater than or equal to 2, let M be a subset of $\mathbb{R}^d$, and let E be a space of real-valued functions defined on M. The object of interest in this paper is the tail behaviour of a separable zero mean Gaussian process $X = \{X(t)\}_{t \in M}$ taking values in the space E, or rather the tail behaviour of a random variable $S = T(X)$

*Corresponding author.
E-mail addresses: koning@few.eur.nl (A.J. Koning), protassov@few.eur.nl (V. Protasov).

doi:10.1016/S0047-259X(03)00059-9

where $T : E \to \mathbb{R}$ is of the type

$$T(f) = \sup_{v \in V} \sqrt{Q_v(f,f)} \tag{1}$$

for every $f \in E$. Here V is some index set, and Q_v is a symmetric bounded bilinear form on E for every $v \in V$. Typically, the random variable S has a quite intricate distribution.

As one may show that any T of the form (1) is sublinear and positive homogeneous, it follows from Theorem 5.2 in [7] that there exists a constant a such that

$$\lim_{y \to \infty} y^{-2} \log P(S > y) = -a/2. \tag{2}$$

Our first aim is to establish methods for the actual computation of the constant a. Our second aim is to construct a random variable R (with a less intricate distribution than S), such that the random variable R

(i) is defined on the same probability space as S, and satisfies

$$P(R \leqslant S) = 1; \tag{3}$$

(ii) has the same tail behaviour as S, in the sense that

$$\lim_{y \to \infty} y^{-2} \log P(R > y) = -a/2, \tag{4}$$

where the constant a is as in (2).

The motivation for the present study comes from the theory of statistical tests, where random variables S emerge as the limit in distribution under the null hypothesis of a sequence of test statistics. Examples will be given shortly.

As the constant a provides a convenient rough description of the limiting distribution of the test statistic at hand, the verification of (2) is a key step in the comparison of statistical tests. In fact, (2) appears as a condition in results for determining approximate Bahadur efficiency (cf. [5]), in results guaranteeing the coincidence of limiting approximate Bahadur efficiency and limiting Pitman efficiency (cf. [27,44]), and in deviation results (cf. [24,35,36]). Deviation results are in turn needed for the computation of Bayes risk efficiency (cf. [38]), intermediate efficiency (cf. [26]) and exact Bahadur efficiency (cf. [5]). Refer to Chapter 1 in [36] and Chapter 10 in [40] for additional information on efficiency concepts.

For a given testing problem each of the efficiency concepts mentioned above may be used to select an "optimal" statistical test. However, when applying the selected test the rough description a is no longer sufficient, and additional precision is needed to determine the critical value (that is, a selected upper percentage point of the test statistic) and/or the attained significance level of the test. In such a situation we resort to the random variable R in order to obtain a more detailed description of tail behaviour of S.

We take a special interest in the situation where the time space and the covariance function both have product structure; that is, we have $M = M_1 \times M_2$ and

$$K((s_1, s_2), (t_1, t_2)) = K_1(s_1, t_1)K_2(s_2, t_2) \quad (5)$$

for every $s_i, t_i \in M_i$ $(i = 1, 2)$. An important example is the situation where M is equal to $[0, 1]^2$, and X coincides with the process $\Gamma = \{\Gamma(t_1, t_2)\}_{0 \leqslant t_1, t_2 \leqslant 1}$, a mean zero Gaussian process with covariance function

$$\mathscr{E}\Gamma(s_1, s_2)\Gamma(t_1, t_2) = \{\min(s_1, t_1) - s_1 t_1\}\{\min(s_2, t_2) - s_2 t_2\} \quad (6)$$

for $0 \leqslant s_1, s_2, t_1, t_2 \leqslant 1$. In literature, the Gaussian process Γ is called the Wiener pillow ([37, p. 137]; inspired by the fact that $\Gamma(t_1, t_2) = 0$ almost surely for all (t_1, t_2) on the boundary of the unit square), the completely tucked Brownian sheet [42, p. 368] or the tied-down Kiefer process [11, p. 320]. We shall refer to Γ as the Brownian pillow. One may view the Brownian pillow as a two-parameter generalization of the Brownian bridge (that is, a one-parameter zero mean Gaussian process $B(t)$ defined on the unit interval [0,1] with covariance function $\mathscr{E}B(s)B(t) = \min(s, t) - st$ for $0 \leqslant s, t \leqslant 1$).

Limiting random variables of the type $T(\Gamma)$ occur in certain nonparametric statistical applications, such as in nonparametric testing of bivariate independence (cf. [6,12–14,18,22]), and nonparametric testing of univariate constancy over time (cf. [11,21]).

Other mean zero Gaussian processes which obey (5) emerge as limiting processes in nonparametric testing of multivariate independence (for instance, the p-variate Hoeffding, Blum, Kiefer and Rosenblatt process with $M = [0, 1]^p$ and covariance function $\prod_{i=1}^{p} \{\min(s_i, t_i) - s_i t_i\}$) and in nonparametric testing of multivariate constancy over time (for instance, the p-variate Gaussian processes with $M = [0, 1]^p$ and covariance function

$$\{F(\min(s_1, t_1), \ldots, \min(s_{p-1}, t_{p-1})) - F(s_1, \ldots, s_{p-1})F(t_1, \ldots, t_{p-1})\}$$
$$\times \{\min(s_p, t_p) - s_p t_p\}$$

of Theorem 2.6.1 in [11, p. 153]).

The structure of the paper is as follows. In Section 2 we first consider the situation in which no structure is imposed on the "time space" M; the results are exemplified using the Brownian bridge. In Section 3 we explore the situation where the time space and the covariance function obey (5); the results are exemplified using the Brownian pillow. In Section 4 we discuss the use of the random variable R in simulating upper percentage points of S. In Section 5 we consider the extension of Proposition 3.2.1, the main result of Section 3, to more general classes of functionals.

2. General Gaussian processes

2.1. Reproducing kernel Hilbert space

Let M be the closure of an open bounded domain in $\mathbb{R}^d$, and let X be a separable zero mean Gaussian process defined on M. Define the covariance function $K : M \times M \to \mathbb{R}$ by $K(s,t) = \mathscr{E}X(s)X(t)$ for $s,t \in M$. As a covariance function is nonnegative definite, there exists a unique Hilbert space $\mathscr{H}$ such that the reproducing property

$$\langle f, K(t,\cdot) \rangle_{\mathscr{H}} = f(t) \quad \text{for every } t \in M,$$

holds for every $f \in \mathscr{H}$ (cf. [1, Eq. (3.9), p. 67]). The Hilbert space H is called the reproducing kernel Hilbert space belonging to X. Refer to [4] for the general theory of reproducing kernels.

If the set M is equipped with a σ-additive measure μ so that the covariance function K belongs to the space $L_2(M \times M, \mu \times \mu)$, then one may describe the reproducing kernel Hilbert space belonging to X by means of the ordered eigenvalues $\{\tau_1 \geqslant \tau_2 \geqslant \cdots \geqslant 0\}$ and the corresponding normalized eigenfunctions of the operator $\mathscr{K}_\mu$ defined by

$$\mathscr{K}_\mu f(t) = \int_M K(s,t) f(s)\, d\mu \quad \text{for every } f \in L_2(M,\mu).$$

It is well known that the operator $\mathscr{K}_\mu$ is self-conjugate and compact (cf. [32]). Hence, by the Hilbert-Schmidt theorem (cf. [30, p. 78]) it has a complete orthonormal system of eigenfunctions $\{h_1, h_2, \dots\}$ and corresponding eigenvalues $\{\tau_1, \tau_2, \dots\}$ that converges to zero. That is, h_j and τ_j solve the integral equation

$$\int_M K(s,t) h_j(s)\, d\mu = \tau_j h_j(t) \tag{7}$$

and h_j and h_k satisfy

$$\int_M h_j(s) h_k(s)\, d\mu = \begin{cases} 1 & \text{for } j = k, \\ 0 & \text{for } j \neq k. \end{cases}$$

As $\mathscr{K}_\mu$ is nonnegative definite due to the fact that $K(s,t)$ is a covariance function, we see that $\tau_1 \geqslant \tau_2 \geqslant \cdots \geqslant 0$. If the sum $\sum_j \tau_j$ is finite (which is the case when the covariance function $K(s,t)$ is continuous and the measure μ is absolutely continuous with respect to Lebesgue measure), then $\mathscr{K}_\mu$ is a trace-class operator, and the kernel $\mathscr{K}$ is represented in the form

$$\mathscr{K}(s,t) = \sum_{j=0}^{\infty} \tau_j h_j(s) h_j(t),$$

where the convergence is uniform on $M \times M$ (see, for instance, [28] for the general theory of trace-class operators).

Lemma 1 follows by Theorem 3.16 in [1, p. 75].

Lemma 1. *The reproducing kernel Hilbert space $\mathscr{H}$ belonging to X is given by*

$$\mathscr{H} = \left\{ f: f = \sum_{j \geqslant 1, \tau_j \neq 0} b_j h_j, \sum_{j \geqslant 1, \tau_j \neq 0} \frac{(b_j)^2}{\tau_j} < \infty \right\},$$

and has scalar product

$$\langle f, g \rangle_{\mathscr{H}} = \sum_{j \geqslant 1, \tau_j \neq 0} \frac{1}{\tau_j} \left\{ \int_M f(s) h_j(s)\, d\mu(s) \right\} \left\{ \int_M g(s) h_j(s)\, d\mu(s) \right\}.$$

Let us remark that the Hilbert space $\mathscr{H}$ is uniquely defined by the kernel K and does not depend on the choice of the absolutely continuous measure μ, although the eigenfunctions $\{h_i\}$ and the eigenvalues $\{\tau_i\}$ may be different for different measures.

2.2. Tail behaviour of a Gaussian process

For the reproducing kernel Hilbert space $\mathscr{H}$ belonging to X consider a positive homogeneous sublinear functional $T: \mathscr{H} \to \mathbb{R}$ (that is $T(f) \geqslant 0, T(\lambda f) = |\lambda| T(f), T(f+g) \leqslant T(f) + T(g)$ for all $f, g \in \mathscr{H}$, $\lambda \in \mathbb{R}$) and define its $\mathscr{H}$-norm by

$$||T||_{\mathscr{H}} = \sup_{f \in \mathcal{O}_{\mathscr{H}}} T(f),$$

where $\mathcal{O}_{\mathscr{H}}$ denotes the unit ball in $\mathscr{H}$. The relevance of this norm for describing the right-hand tail of the distribution of $S = T(X)$ is shown by the next inequality appearing in the proof of Theorem 5.2 in [7].

Inequality 2.1 (Inequality of Borell). *Let T be a positive homogeneous sublinear functional on $\mathscr{H}$. Suppose there exists x_0 such that $P(S \leqslant x_0)$ is positive, and let ξ satisfy $P(Z \leqslant \xi) \leqslant P(S \leqslant x_0)$, where Z is a standard normal random variable. Then*

$$P(S \geqslant x) \leqslant P(Z > \xi + ||T||_{\mathscr{H}}^{-1}(x - x_0))$$

for every $x > x_0$.

The Inequality of Borell implicitly provides an exponential bound for $P(S \geqslant x)$, as readily follows from the next inequality (cf. [41, p. 850]).

Inequality 2.2 (Mill's ratio). *Let Z be a standard normal random variable. Then, for all $y > 0$,*

$$(y^{-1} - y^{-3}) \frac{1}{\sqrt{2\pi}} e^{-y^2/2} \leqslant P(Z > y) \leqslant y^{-1} \frac{1}{\sqrt{2\pi}} e^{-y^2/2}.$$

If there exists a constant $c>0$ such that

$$T(f) \leqslant c \max\left\{\sup_{t\in M} |f|, 1\right\} \quad \text{for every } f \in \mathscr{H} \tag{8}$$

(as is usually the case in statistical applications), then the existence of x_0 such that $P(S \leqslant x_0)$ is positive can be verified with the aid of the following version of Borell's inequality (cf. [39]).

Inequality 2.3 (Inequality of Borell, supremum version). *If* $\mathscr{E}\sup_{t\in M} X(t) < \infty$, *then*

$$P\left(\sup_t X(t) \geqslant x\right) \leqslant 2\exp\left\{-\frac{(x-\mathscr{E}\sup_{t\in M} X(t))^2}{2\sup_{t\in M} K(t,t)}\right\}$$

for every $x \geqslant \mathscr{E}\sup_{t\in M} X(t)$.

If Q is a bounded bilinear form, then the functional T defined by $T(f) = \sqrt{Q(f,f)}$ for every $f \in \mathscr{H}$ satisfies (8) (cf. Definition 13.2 in [20, p. 89]).

For Gaussian processes the boundedness of the supremum is intimately related to sample path continuity (cf. Section III.1 in [1, p. 62]).

2.3. Tail behaviour of supremum and quadratic tests

Consider a statistical problem, where at stage n it is natural to base statistical tests on a "monitoring process" X_n, which under the null hypothesis converges in distribution to X as n tends to infinity. As an example, one may think of the goodness-of-fit problem, the independence problem and the change-point problem. An appropriate monitoring process in the goodness-of-fit problem is the multivariate empirical process, which converges under the null hypothesis to the tied-down Brownian motion (cf. [15,17]). An appropriate monitoring process in the independence and change-point problems is the Hoeffding, Blum, Kiefer, Rosenblatt multivariate empirical process, which converges under the null hypothesis to Brownian pillow type processes [6,9,10,22].

In such statistical problems, obvious tests are supremum and quadratic tests derived from the monitoring process X_n. Supremum tests reject when the supremum test statistic $\sup_{t\in M} |X_n(t)|$ becomes large. Under the null hypothesis, this statistic converges in distribution to $T(X)$, where the corresponding functional $T(f) = \sup_{t\in M} |f(t)|$ is of form (1) with $V = M$ and $Q_v(f,g) = f(v)g(v)$. The Kolmogorov test is an example of a supremum test.

Let Q be a symmetric bounded bilinear form. The quadratic test corresponding to Q rejects when the "quadratic" test statistic $\sqrt{Q(X_n, X_n)}$ becomes large. Under the null hypothesis, the quadratic test statistic converges in distribution to $T(X)$, where T is of form (1) with V equal to a singleton $\{v_0\}$ and Q_{v_0} equal to Q. The Cramér–von Mises and Anderson–Darling tests are examples of quadratic tests.

Our study of random variables $T(X)$ where T is of the form (1) starts by observing that for every $v \in V$ there exists a unique bounded linear operator $\mathscr{A}_v$

defined on $\mathscr{H}$ such that

$$Q_v(f,g) = \langle f, \mathscr{A}_v g \rangle_{\mathscr{H}}$$

for every $f,g \in \mathscr{H}$ (cf. Theorem 13.5b in [20, p. 92]). Observe that $\mathscr{A}_v$ depends both on the choice of Q_v and [via the norm $\langle \cdot, \cdot \rangle_{\mathscr{H}}$] on the covariance function K of X. Assume that Q_v satisfies the following condition.

Condition 1. For each $v \in V$ the operator $\mathscr{A}_v$ associated to Q_v has a complete orthonormal system of eigenfunctions $\{\phi_{1,v}, \phi_{2,v}, \dots\}$. Moreover, there exists $w \in V$ such that $\sup_{v \in V} \lambda_{1,v} = \lambda_{1,w}$, where $\lambda_{1,v}$ denotes the largest eigenvalue of $\mathscr{A}_v$.

Let us recall that in general a self-conjugate positively definite operator in a Hilbert space may have no eigenvalues at all [30, p. 273], so Condition 1 is indeed quite restrictive. Nevertheless, in most statistical applications this condition does hold.

The second part of Condition 1 is fulfilled when, for instance, $\lambda_{1,v}$ is a continuous function of v, and V is compact.

Lemma 2. *If Condition* 1 *holds*, *then*

(i) $||T||_{\mathscr{H}}$ *is equal to* $T(\phi_{1,w}) = \sqrt{\lambda_{1,w}}$;
(ii) $T(X)$ *is larger than or equal to* $||T||_{\mathscr{H}} \cdot |Z|$ *with probability* 1, *where* $Z = \langle X, \phi_{1,w} \rangle_{\mathscr{H}}$ *is a standard normal random variable.*

Denote the random variable $||T||_{\mathscr{H}} \cdot |Z|$ by R. It follows from Inequality 2.2 that (4) holds with

$$a = ||T||_{\mathscr{H}}^{-2}. \tag{9}$$

As Inequality 2.1 implies

$$\limsup_{y \to +\infty} (y)^{-2} \log P(T(X) > y) \leqslant -a/2$$

with a given by (9), it immediately follows from Lemma 2 that the random variables $T(X)$ and R have similar tail behaviour, in the sense that

$$\lim_{y \to +\infty} (y)^{-2} \log P(T(X) > y) = \lim_{y \to +\infty} (y)^{-2} \log P(R > y). \tag{10}$$

In particular, we obtain that (2) holds with a given by (9).

As noted in the introduction, (2) is directly relevant for the computation of the approximate Bahadur efficiency of the test based on $T(X_n)$. Moreover, in combination with a KMT-type approximation for X_n, (2) implies a deviation result

$$\lim_{n \to +\infty} (y_n)^{-2} \log P(T(X_n) > y_n) = -a/2 \tag{11}$$

for sequences y_n that tend to infinity at a sufficiently slow rate as $n \to \infty$. A KMT-type approximation is a strong approximation governed by an exponential inequality, as the ones given in [33] for the partial sum process and the empirical process.

The quality of the KMT-type approximation determines the maximal rate of y_n allowed in (11) (cf. [23,24,34,35]). Special deviation results are:

- Chernoff-type deviation results, which allow sequences y_n up to $O(n^{1/2})$, and are relevant for the computation of exact Bahadur efficiency [5];
- Cramér-type deviation results, which allow sequences y_n up to $o(n^{1/6})$, and are relevant for the computation of intermediate efficiency [26];
- moderate deviation results, which only allow sequences y_n up to $O((\log n)^{1/2})$, and are relevant for the computation of Bayes risk efficiency [38] and weak intermediate efficiency [26].

For some of the more popular functionals (for instance, the functionals T_{CvM} and T_{AD} introduced at the end of this section), we have $T(X)$ is equal in distribution to $\sqrt{\sum_{j=1}^{\infty} \lambda_{j,w} Z_j^2}$, where $Z_1, Z_2, \ldots$ is a sequence of independent standard normal random variables. For such functionals , Lemma 2 seems to be related to Lemma 2.4 in [19], which states that

$$\lim_{x \to +\infty} (x)^{-1} \log P\left(\sum_{j=1}^{\infty} \lambda_{j,w}(Z_j^2 - 1) > x \right) = -(2\lambda_{j,w})^{-1}$$

if $\sum_{j=1}^{\infty} \lambda_{j,w}^2$ is finite.

Remark that (10) suggests that for small significance levels the critical value of the test statistic $T(X)$ may be approximated by the corresponding quantile of the random variable R. However, such an approach is not recommended, since it would lead to a anti-conservative approximate test.

2.4. Application to the Brownian bridge

Recall that a Brownian bridge is a zero mean Gaussian process defined on the unit interval, with covariance function $K_B(s,t) = \min(s,t) - st$. For a differentiable function $f : [0,1] \to \mathbb{R}$, let f' denote the derivative of f.

By differentiating both sides of (7) twice (here and below the integrals are computed with respect to the usual Lebesque measure), it follows that the eigenvalues and eigenfunctions belonging to the Brownian bridge are solutions to the differential equation

$$\tau h''(t) = -h(t) \tag{12}$$

under the boundary condition $h(0) = h(1) = 0$. Hence, these eigenvalues and eigenfunctions are given by

$$\tau_j = (j\pi)^{-2}, \quad h_j(t) = \sqrt{2}\sin(j\pi t)$$

(cf. Proposition 5.3.1 in [41, p. 213]). Thus, we may invoke Lemma 1 to describe the reproducing kernel Hilbert space $\mathscr{H}_B$ belonging to the Brownian bridge B.

However, for functions f, g in $\mathscr{H}_B$ which do not have a simple and clear relation with the Brownian bridge eigenfunctions h_j the computation of the scalar product may well be intricate. Lemma 3 provides an alternative representation of the scalar product in $\mathscr{H}_B$, which often is more convenient.

Lemma 3. *The Hilbert space $\mathscr{H}_B$ corresponding to the kernel of the Brownian bridge B is given by*

$$\mathscr{H}_B = \{f : [0,1] \to \mathbb{R},\ f' \in L_2[0,1],\ f(0) = f(1) = 0\},$$

and has scalar product

$$\langle f, g \rangle_{\mathscr{H}_B} = \int_0^1 f'(t) g'(t)\, dt.$$

If f and g belong to $\mathscr{H}_B$, and g is twice differentiable, then it follows by integration by parts that we may write

$$\langle f, g \rangle_{\mathscr{H}_B} = -\int_0^1 f(t) g''(t)\, dt.$$

Example 1 (The Kolmogorov functional). Consider the functional T_{Kol} defined by

$$T_{\mathrm{Kol}}(f) = \sup_{t \in [0,1]} |f(t)|.$$

Observe that T_{Kol} satisfies (1) with $V = [0,1]$ and $Q_v = f(v)g(v)$. The corresponding operator is $\mathscr{A}_v f(t) = f(v) \cdot \phi_{1,v}(t)$, where

$$\phi_{1,v}(t) = \begin{cases} t\sqrt{\frac{1-v}{v}} & \text{if } t \leqslant v, \\ (1-t)\sqrt{\frac{v}{1-v}} & \text{if } t \geqslant v. \end{cases}$$

It is seen easily that $\phi_{1,v}$ is an eigenfunction of the operator $\mathscr{A}_v$ corresponding to the largest eigenvalue $\lambda_{1,v} = v(1-v)$, all other eigenvalues $\lambda_{j,v}$, $j \geqslant 2$ are zeros. Thus Condition 1 holds, and the maximal eigenvalue $\lambda_{1,w} = \frac{1}{4}$ is attained for $w = \frac{1}{2}$. It now follows by Lemma 2(i) that in $\mathscr{H}_B$ the norm of T_{Kol} is equal to 2^{-1}. Moreover, for $w = 1/2$ we have

$$\langle B, \phi_{1,w} \rangle_{\mathscr{H}_B} = \int_{[0,1]} \phi'_{1,w}(t)\, dB(t) = \int_{[0,1/2]} dB(t) - \int_{[1/2,1]} dB(t) = 2B(1/2),$$

and hence it follows by Lemma 2(ii) that $T_{\mathrm{Kol}}(B)$ is larger than or equal to $||T_{\mathrm{Kol}}||_{\mathscr{H}_B} \cdot |Z| = |Z/2|$, where $Z = 2B(1/2)$ is a standard normal random variable.

The exact distribution of $T_{\mathrm{Kol}}(B)$ is given in [31].

Example 2 (The Cramér–von Mises functional). Consider the functional T_{CvM} defined by

$$T_{\mathrm{CvM}}(f) = \left\{ \int_0^1 (f(s))^2\, ds \right\}^{1/2}.$$

Observe that T_{CvM} satisfies (1) with $V = \{v_0\}$ and $Q_{v_0} = \int_0^1 f(s)g(s)\, ds$. Define $\mathscr{A}$ by $\mathscr{A}f(s) = \int_0^1 K(s,t)f(t)\, dt$, where $K(s,t) = \min(s,t) - st$ is our reproducing kernel. It is easy to see that this operator takes the space $\mathscr{H}_B$ to the space $\mathscr{H}'_B = \{f \in \mathscr{H}_B \,|\, f'' \in \mathscr{H}_B\}$. As $\mathscr{A}^{-1}F = -F''$ for every $F \in \mathscr{H}'_B$, the eigenfunctions of $\mathscr{A}$ satisfy $-f'' = \lambda^{-1}f$ (cf. with (12)), and hence the eigenvalues and the normalized eigenfunctions are given by

$$\lambda_j = (j\pi)^{-2}, \quad \phi_j(t) = (j\pi)^{-1}\sqrt{2}\sin(j\pi t).$$

In particular, we have $\lambda_1 = \pi^{-2}$ and $\phi_j(t) = \pi^{-1}\sqrt{2}\sin(\pi t)$. Since the operator $\mathscr{A}$ is compact and self-conjugate, therefore, using again the Hilbert-Schmidt theorem, we conclude that $\mathscr{A}$ is diagonalized i.e., $\{\phi_j\}$ is indeed an orthonormal basis in $\mathscr{H}_B$. Thus, Condition 1 holds. It follows by Lemma 2(i) that in $\mathscr{H}_B$ the norm of T_{CvM} is equal to π^{-1}. Hence, by Lemma 2(ii) $T_{\mathrm{CvM}}(B)$ is larger than or equal to $||T_{\mathrm{CvM}}||_{\mathscr{H}_B} \cdot |Z| = |\pi^{-1}Z|$, where

$$Z = \langle B, \phi_1 \rangle_{\mathscr{H}_B} = -\int_M B(s)\phi_1''(s)\, ds = \pi\sqrt{2}\int_M B(s)\sin(\pi s)\, ds$$

is a standard normal random variable. The random variable $\pi^{-1}Z$ coincides with the limit in distribution of the "first component" of the Cramér–von Mises test statistic (cf. [16]).

The exact distribution of $T_{\mathrm{CvM}}(B)$ (or rather $\{T_{\mathrm{CvM}}(B)\}^2$) is described in [2].

Example 3 (The Andersen–Darling functional). Consider the functional T_{AD} defined by

$$T_{\mathrm{AD}}(f) = \left\{ \int_0^1 \frac{(f(s))^2}{s(1-s)}\, ds \right\}^{1/2}.$$

For this functional we have $V = \{v_0\}$ and $Q_{v_0}(f,g) = \int_0^1 (f(s)g(s)/s(1-s))\, ds$.

Recall that $K_B(s,t) = \min(s,t) - st$. Since the operator $\mathscr{A}$ defined by $\mathscr{A}f(t) = \int_0^1 (K_B(s,t)/s(1-s))f(s)\, ds$ satisfies $\mathscr{A}^{-1}F(t) = -t(1-t)F''(t)$, it follow that the eigenfunctions and the eigenvalues are found from the equation $\lambda^{-1}F(t) = -t(1-t)F''(t)$. The solutions of this equation are $F_j(t) = t(1-t)L_{j-1}(2t-1)$,

$\lambda_j = \frac{1}{j(j+1)}$, $j = 1, 2, \ldots$, where L_j is the Legendre polynomial of degree j (cf. [43, p. 324]). Since the operator $\mathscr{A}$ is self-conjugate, it follows that these solutions form an orthogonal system in $\mathscr{H}_B$. The system of Legendre polynomials is complete in $L_2[-1, 1]$, consequently the system $\{F_j\}$ is complete in $\mathscr{H}_B$. Thus, after normalization we obtain the complete orthonormal system of eigenfunctions

$$\phi_j(t) = \sqrt{2\frac{2j+1}{j(j+1)}}\, t(1-t)L_{j-1}(2t-1)$$

and corresponding eigenvalues $\lambda_j = \frac{1}{j(j+1)}$, $j \geqslant 1$ (see also [16, p. 303]). Hence Condition 1 is fulfilled. In particular, we have $\phi_1(t) = \sqrt{3}t(1-t)$ and $\lambda_1 = 1/2$. It follows by Lemma 2(i) that in $\mathscr{H}_B$ the norm of T_{AD} is equal to $2^{-1/2}$. It follows by Lemma 2(ii) that $T_{\mathrm{AD}}(B)$ is larger than or equal to $||T_{\mathrm{AD}}||_{\mathscr{H}_B} \cdot |Z| = |2^{-1/2}Z|$, where

$$Z = \langle B, \phi_1 \rangle_{\mathscr{H}_B} = -\int_0^1 B(s)\phi_1''(s)\,ds = 2\sqrt{3}\int_0^1 B(s)\,ds$$

by Lemma 1. The random variable $2^{-1/2}Z$ coincides with the limit in distribution of the "first component" of the Anderson-Darling test statistic (cf. [16]).

The exact distribution of $T_{\mathrm{AD}}(B)$ (or rather $\{T_{\mathrm{AD}}(B)\}^2$) is described in [3].

3. Covariance functions with product structure

3.1. Reproducing kernel Hilbert space

In this section we consider the situation where the covariance function of X obeys the product structure as given by (5). As an example, one may think of the limit in distribution of the Hoeffding, Blum, Kiefer, Rosenblatt p-variate empirical process, which is a mean zero Gaussian process with covariance function $\prod_{i=1}^p \{\min(s_i, t_i) - s_i t_i\}$ (see Section 3 in [6]). Observe that for $p = 2$, this process coincides with the Brownian pillow.

Although we concentrate on the product structure (5), our results have direct implications for the situation where

$$K((s_1, s_2, \ldots, s_p), (t_1, t_2, \ldots, t_p)) = \prod_{i=1}^p K_i(s_i, t_i)$$

for every $s_i, t_i \in M_i$ $(i = 1, 2, \ldots, p)$.

Lemma 4. *The reproducing kernel Hilbert space* $\mathscr{H}$ *of* X *is equal to the tensor product* $\mathscr{H}_1 \circ \mathscr{H}_2$, *where* $\mathscr{H}_i$ *denotes the reproducing kernel Hilbert space corresponding to* $K_i(x_i, y_i)$ $(i = 1, 2)$.

The tensor product $\mathscr{H}_1 \circ \mathscr{H}_2$ is the Hilbert space with basis $\{h_{1j} \circ h_{2k}\}$, where $\{h_{ij}\}$ is an orthonormal basis of $\mathscr{H}_i$. For any $f_1 = \sum_j a_{1j} h_{1j} \in \mathscr{H}_1$ and $f_2 = \sum_k a_{2k} h_{2k} \in \mathscr{H}_2$ we have

$$f_1 \circ f_2 = \sum_j \sum_k a_{1j} a_{2k} \cdot h_{1j} \circ h_{2k}.$$

The scalar product in $\mathscr{H}_1 \circ \mathscr{H}_2$ is defined as

$$\left\langle \sum_{j_1} \sum_{k_1} a_{j_1 k_1} \cdot h_{1 j_1} \circ h_{2 k_1}, \sum_{j_2} \sum_{k_2} b_{j_2 2 k_2} \cdot h_{1 j_2} \circ h_{2 k_2} \right\rangle_{\mathscr{H}_1 \circ \mathscr{H}_2} = \sum_{j,k} a_{jk} b_{jk}.$$

It is clear that for any $f_1, g_1 \in \mathscr{H}_1$ and $f_2, g_2 \in \mathscr{H}_2$ we have

$$\langle f_1 \circ f_2, g_1 \circ g_2 \rangle_{\mathscr{H}_1 \circ \mathscr{H}_2} = \langle f_1, g_1 \rangle_{\mathscr{H}_1} \langle f_2, g_2 \rangle_{\mathscr{H}_2},$$

and in particular

$$\langle h_{1 j_1} \circ h_{2 k_1}, h_{1 j_2} \circ h_{2 k_2} \rangle_{\mathscr{H}_1 \circ \mathscr{H}_2} = \delta_{j_1, j_2} \delta_{k_1, k_2}.$$

It is seen easily that the tensor product of Hilbert spaces does not depend on the choice of the orthonormal bases in them.

3.2. Tail behaviour of supremum and quadratic tests

For $i = 1, 2$, let $\mathscr{H}_i$ be a Hilbert space. Let Q_i be a symmetric and bounded (not necessarily positively semidefinite) bilinear form on $\mathscr{H}_i$; that is, there exists a positive constant c such that

$$|Q_i(f_i, g_i)| = |Q_i(g_i, f_i)| \leqslant c \|f_i\|_{\mathscr{H}_i} \cdot \|g_i\|_{\mathscr{H}_i}$$

for every $f_i, g_i \in \mathscr{H}_i$. Then one can define a bilinear form $Q_1 \circ Q_2$ on the tensor product $\mathscr{H}_1 \circ \mathscr{H}_2$ as follows: for the elements of basis we set $Q_1 \circ Q_2(h_{1 j_1} \circ h_{2 j_2}, h_{1 k_1} \circ h_{2 k_2}) = Q_1(h_{1 j_1}, h_{1 k_1}) Q_2(2 h_{j_2}, h_{2 k_2})$, then extend this form onto $\mathscr{H}_1 \circ \mathscr{H}_2$ by bilinearity. The tensor product of bilinear forms does not depend on the choice of bases and possesses the property $Q_1 \circ Q_2(f_1 \circ f_2, g_1 \circ g_2) = Q_1(f_1, g_1) Q_2(f_2, g_2)$ for all $f_i, g_i \in \mathscr{H}_i$.

Proposition 3.2.1. *Let $\mathscr{H}_i$ be a Hilbert space, Q_i be a symmetric and bounded bilinear form on $\mathscr{H}_i$, $i = 1, 2$. Then*

$$\sup_{f \in \mathscr{H}_1 \circ \mathscr{H}_2} |Q_1 \circ Q_2(f)| = \sup_{f_1 \in \mathscr{H}_1} |Q_1(f_1)| \sup_{f_2 \in \mathscr{H}_2} |Q_2(f_2)|. \tag{13}$$

If the suprema on the RHS of (13) *are respectively attained for $f_1 = f_1^*$ and $f_2 = f_2^*$, then the supremum on the LHS of* (13) *is attained for $f = f_1^* \circ f_2^*$.*

Corollary 3.1. *Let the functional T_i be defined by $T_i(f_i) = \{\sup_{v_i \in V_i} |Q_{i,v_i}(f_i)|\}^{1/2}$ for $f_i \in \mathscr{H}_i$, where V_i is some index set and Q_{i,v_i} is a symmetric bounded bilinear form* *satisfying Condition* 1. *Then the functional* $T = T_1 \circ T_2$ *defined by*

$$T(f) = \left\{ \sup_{v_1 \in V_1} \sup_{v_2 \in V_2} |Q_{1,v_1} \circ Q_{2,v_2}(f)| \right\}^{1/2}$$

for $f \in \mathscr{H}_1 \circ \mathscr{H}_2$ *satisfies* $||T||_{\mathscr{H}_1 \circ \mathscr{H}_2} = ||T_1||_{\mathscr{H}_1} \cdot ||T_2||_{\mathscr{H}_2}$.

Lemma 5. *Let* T *be as in Corollary* 3.1, *let* ϕ_{i1,w_i} *denote the eigenfunction corresponding to the largest eigenvalue of* Q_{i,w_i}, *and define the process* X_2 *by*

$$X_2(t_2) = \langle X(\cdot, t_2), \phi_{11,w_1} \rangle_{\mathscr{H}_1}$$

for $t_2 \in M_2$. *Then* X_2 *is a zero mean Gaussian process* X_2 *with covariance function* $K_2(s_2, t_2)$. *Moreover*, $T(X)$ *is larger than or equal to* $||T_1||_{\mathscr{H}_1} \cdot T_2(X_2)$ *with probability* 1.

Remark that Corollary 3.1 implies that the random variables $T(X)$ and $||T_1||_{\mathscr{H}_1} \cdot T_2(X_2)$ have similar tail behaviour, in the sense that

$$\lim_{y \to +\infty} (y)^{-2} \log P(T(X) > y) = \lim_{y \to +\infty} (y)^{-2} \log P(||T_1||_{\mathscr{H}_1} \cdot T_2(X_2) > y). \tag{14}$$

Applying Lemma 2(ii) to X_2 yields that there exists a standard normal random variable Z such that

$$T(X) \geqslant ||T_1||_{\mathscr{H}_1} \cdot T_2(X_2) \geqslant ||T_1||_{\mathscr{H}_1} \cdot ||T_2||_{\mathscr{H}_2} \cdot |Z|$$

with probability 1. This indicates that Lemma 5 may well yield better results than Lemma 2(ii) when applied to a zero mean Gaussian process with product structure in the covariance function.

By a symmetry argument, it follows that under the conditions of Lemma 5 we also have that $T(X)$ is larger than or equal to $||T_2||_{\mathscr{H}_2} \cdot T_1(X_1)$ with probability 1, where the process X_1 is defined by

$$X_1(t_1) = \langle X(t_1, \cdot), \phi_{21,w_2} \rangle_{\mathscr{H}_2}.$$

3.3. Application to the Brownian pillow

As K_Γ satisfies (5) with $K_i = K_B$, we have $\mathscr{H}_\Gamma = \mathscr{H}_B \circ \mathscr{H}_B$. Lemma 6 presents an alternative representation of the reproducing kernel Hilbert space belonging to Γ. For a function $f : [0,1]^2 \to \mathbb{R}$ which is differentiable in both components, let f_{12} denote the partial derivative of f obtained by differentiating with respect to both components.

Lemma 6. *The Hilbert space* $\mathscr{H}_\Gamma$ *corresponding to the kernel of the Brownian pillow* Γ *is given by*

$$\mathscr{H}_\Gamma = \{f : [0,1]^2 \to \mathbb{R}, f_{12} \in L_2[0,1], f(t_1, t_2) = 0 \textit{ on the boundary of } [0,1]^2\},$$

and has scalar product

$$\langle f,g\rangle_{\mathscr{H}_\Gamma} = \int_0^1 \int_0^1 f_{12}(t_1,t_2)g_{12}(t_1,t_2)\,dt_1\,dt_2.$$

Example 4 (An extension of the Kolomogorov functional). Let $T_{\rm Kol}$ be as in the previous section, recall that $||T_{\rm Kol}||_{\mathscr{H}_B} = 1/2$.

- For the functional $T = T_{\rm Kol}\circ T_{\rm Kol}$ defined by

  $$T_{\rm Kol}\circ T_{\rm Kol}(f) = \sup_{t_1,t_2\in[0,1]} |f(t_1,t_2)| \quad \text{for every } f\in\mathscr{H}_\Gamma.$$

  Corollary 3.1 yields $||T||_{\mathscr{H}_\Gamma} = ||T_{\rm Kol}||^2_{\mathscr{H}_B} = 1/4$. By Lemma 2(ii) there exists a standard normal random variable Z such that $T(\Gamma)\geqslant|Z|/4$. By Lemma 5 we have $T(\Gamma)\geqslant||T_{\rm Kol}||_{\mathscr{H}_B}T_{\rm Kol}(B_{\rm Kol}) = T_{\rm Kol}(B_{\rm Kol})/2$ with probability 1, where the Brownian bridge $B_{\rm Kol}$ is defined by

  $$B_{\rm Kol}(t_2) = 2\Gamma(1/2,t_2) \quad \text{for every } t_2\in[0,1].$$

  To the author's knowledge, the distribution of $T(\Gamma)$ is not known. The only result found in literature with respect to this distribution is the upper bound in [29] (cf. [6]): there exist unspecified positive constants c_1, c_2 such that

  $$P(T(\Gamma)>y)\leqslant c_1\exp\{-c_2y^2\}.$$

  Observe that this upper bound follows from the Inequality of Borell. In fact, we may take c_2 equal to $1/2||T||^2_{\mathscr{H}_\Gamma} = 8$.

Example 5 (Extensions of the Cramér–von Mises functional). Let $T_{\rm CvM}$ be as in the previous section, recall that $||T_{\rm CvM}||_{\mathscr{H}_B} = \pi^{-1}$.

- For the functional $T = T_{\rm CvM}\circ T_{\rm CvM}$ defined by

  $$T_{\rm CvM}\circ T_{\rm CvM}(f) = \left\{\int_0^1 \int_0^1 f^2(t_1,t_2)\,dt_1\,dt_2\right\}^{1/2} \quad \text{for every } f\in\mathscr{H}_\Gamma,$$

  Corollary 3.1 yields $||T||_{\mathscr{H}_\Gamma} = ||T_{\rm CvM}||^2_{\mathscr{H}_B} = \pi^{-2}$. By Lemma 2(ii) there exists a standard normal random variable Z such that $T(\Gamma)\geqslant|Z|/\pi^2$. By Lemma 5 we have $T(\Gamma)\geqslant||T_{\rm CvM}||_{\mathscr{H}_B}\cdot T_{\rm CvM}(B_{\rm CvM}) = T_{\rm CvM}(B_{\rm CvM})/\pi$ with probability 1, where $B_{\rm CvM}$ is the Brownian bridge defined by

  $$B_{\rm CvM}(t_2) = \pi\sqrt{2}\int_0^1 \Gamma(t_1,t_2)\sin(\pi t_1)\,dt_1 \quad \text{for every } t_2\in[0,1].$$

  The distribution of $T(\Gamma)$ is tabulated in [6].

- For the functional $T = T_{\rm Kol}\circ T_{\rm CvM}$ defined by

$$T_{\rm Kol}\circ T_{\rm CvM}(f) = \sup_{t_1\in[0,1]}\left\{\int_0^1 f^2(t_1,t_2)\,dt_2\right\}^{1/2} \quad \text{for every } f\in\mathscr{H}_\Gamma,$$

  Corollary 3.1 yields $||T||_{\mathscr{H}_\Gamma} = ||T_{\rm Kol}||_{\mathscr{H}_B}\cdot||T_{\rm CvM}||_{\mathscr{H}_B} = 1/2\pi$. By Lemma 2(ii) there exists a standard normal random variable Z such that $T(\Gamma)\geqslant|Z|/2\pi$. By Lemma 5 we have $T(\Gamma)\geqslant||T_{\rm Kol}||_{\mathscr{H}_B}T_{\rm CvM}(B_{\rm Kol}) = T_{\rm CvM}(B_{\rm Kol})/2$ and $T(\Gamma)\geqslant||T_{\rm CvM}||_{\mathscr{H}_B}.\ T_{\rm Kol}(B'_{\rm CvM}) = T_{\rm Kol}(B'_{\rm CvM})/\pi$ with probability 1; here $B'_{\rm CvM}$ is the Brownian bridge defined by

$$B'_{\rm CvM}(t_1) = \pi\sqrt{2}\int_0^1 \Gamma(t_1,t_2)\sin(\pi t_2)\,dt_2 \quad \text{for every } t_1\in[0,1].$$

Example 6 (Extensions of the Anderson–Darling functional). Let $T_{\rm AD}$ be as in the previous section, recall that $||T_{\rm AD}||_{\mathscr{H}_B} = 2^{-1/2}$.

- For the functional $T = T_{\rm AD}\circ T_{\rm AD}$ defined by

$$T_{\rm AD}\circ T_{\rm AD}(f) = \left\{\int_0^1\int_0^1 \frac{f^2(t_1,t_2)}{t_1(1-t_1)t_2(1-t_2)}\,dt_1\,dt_2\right\}^{1/2} \quad \text{for every } f\in\mathscr{H}_\Gamma,$$

  Corollary 3.1 yields $||T||_{\mathscr{H}_\Gamma} = ||T_{\rm AD}||^2_{\mathscr{H}_B} = 1/2$. By Lemma 2(ii) there exists a standard normal random variable Z such that $T(\Gamma)\geqslant|Z|/2$. By Lemma 5 we have $T(\Gamma)\geqslant||T_{\rm AD}||_{\mathscr{H}_B}\cdot T_{\rm AD}(B_{\rm AD}) = T_{\rm AD}(B_{\rm AD})/\sqrt{\pi}$ with probability 1, where $B_{\rm AD}$ is the Brownian bridge defined by

$$B_{\rm AD}(t_2) = \sqrt{3}\int_0^1 \Gamma(t_1,t_2)\,dt_1 \quad \text{for every } t_2\in[0,1].$$

- For the functional $T = T_{\rm Kol}\circ T_{\rm AD}$ defined by

$$T_{\rm Kol}\circ T_{\rm AD}(f) = \sup_{t_1\in[0,1]}\left\{\int_0^1 \frac{f^2(t_1,t_2)}{t_2(1-t_2)}\,dt_2\right\}^{1/2} \quad \text{for every } f\in\mathscr{H}_\Gamma,$$

  Corollary 3.1 yields $||T||_{\mathscr{H}_\Gamma} = ||T_{\rm Kol}||_{\mathscr{H}_B}\cdot||T_{\rm AD}||_{\mathscr{H}_B} = 2^{-3/2}$. By Lemma 2(ii) there exists a standard normal random variable Z such that $T(\Gamma)\geqslant|Z|/2\sqrt{2}$. By Lemma 5 we have $T(\Gamma)\geqslant||T_{\rm Kol}||_{\mathscr{H}_B}T_{\rm AD}(B_{\rm Kol}) = T_{\rm AD}(B_{\rm Kol})/2$ and $T(\Gamma)\geqslant||T_{\rm AD}||_{\mathscr{H}_B}\cdot T_{\rm Kol}(B'_{\rm AD}) = T_{\rm Kol}(B'_{\rm AD})/\sqrt{2}$ with probability 1; here $B'_{\rm AD}$ is the Brownian bridge defined by

$$B_{\rm AD}(t_1) = \sqrt{3}\int_0^1 \Gamma(t_1,t_2)\,dt_2 \quad \text{for every } t_1\in[0,1].$$

- For the functional $T = T_{\rm CvM}\circ T_{\rm AD}$ defined by

$$T_{\rm CvM}\circ T_{\rm AD}(f) = \left\{\int_0^1\int_0^1 \frac{f^2(t_1,t_2)}{t_2(1-t_2)}\,dt_1\,dt_2\right\}^{1/2} \quad \text{for every } f\in\mathscr{H}_\Gamma,$$

Corollary 3.1 yields $||T||_{\mathscr{H}_\Gamma} = ||T_{\mathrm{CvM}}||_{\mathscr{H}_B} \cdot ||T_{\mathrm{AD}}||_{\mathscr{H}_B} = (\pi\sqrt{2})^{-1}$. By Lemma 2(ii) there exists a standard normal random variable Z such that $T(\Gamma) \geqslant |Z|/\pi\sqrt{2}$. By Lemma 5 yields that $T(\Gamma) \geqslant ||T_{\mathrm{CvM}}||_{\mathscr{H}_B} T_{\mathrm{AD}}(B_{\mathrm{Kol}}) = T_{\mathrm{AD}}(B_{\mathrm{Kol}})/\pi$ and $T(\Gamma) \geqslant ||T_{\mathrm{AD}}||_{\mathscr{H}_B} \cdot T_{\mathrm{CvM}}(B'_{\mathrm{AD}}) = T_{\mathrm{CvM}}(B'_{\mathrm{AD}})/\sqrt{2}$ with probability 1.

4. Refining the results by simulation

Consider a random variable of interest S and a reference variable R satisfying (2)–(4). As noted before, the direct use of the distribution of R as an approximation to the distribution of S should be avoided since it leads to an anti-conservative test. However, if the distribution of R is known, then we may employ simulation methods (using R as a "control variate" for S) to approximate the tail distribution of S. The preceding results may act as guidelines for the statistical analysis of the simulation results. In this section, we illustrate this approach by applying it to statistics of the form $S = T_1 \circ T_2(\Gamma)$, where T_1 and T_2 are either T_{Kol}, T_{CvM} or T_{AD}, and Γ is a Brownian pillow. We remark that the distributions of $T_{\mathrm{Kol}}(B)$, $T_{\mathrm{CvM}}(B)$ and $T_{\mathrm{AD}}(B)$ have been tabulated ([31], [2,3], selected upper percentage points are given in Table 1).

In our simulation study we performed 10.000 simulations. In each simulation generated the Brownian pillow on a 1.000 × 1.000 grid, and computed $S = T_1 \circ T_2(\Gamma)$ and $R = |T_1|_B \circ T_2(B_1)$. Thus, we obtained 10.000 independent copies (S_j, R_j) of (S, R). Let $S^{(1)} \leqslant S^{(2)} \leqslant \cdots \leqslant S^{(10\,000)}$ and $R^{(1)} \leqslant R^{(2)} \leqslant \cdots \leqslant R^{(10\,000)}$ denote the ordered versions of the random samples $S_1, S_2, \ldots, S_{10000}$ and $R_1, R_2, \ldots, R_{10000}$, respectively. We shall refer to $S^{(i)}$ and $R^{(i)}$ as the ith order statistics of S and R, respectively. Observe that $S \geqslant R$ with probability 1 implies $S^{(i)} \geqslant R^{(i)}$ with probability 1.

As we were interested in the tail behaviour of S, we investigated the relation between $S^{(i)}$ and $R^{(i)}$ for $i = 9001, \ldots, 10\,000$ (that is, the upper ten percent of the order statistics) by exploratory statistical methods. For all statistics S under consideration, we found that plots of $\ln(S^{(i)} - R^{(i)})$ versus $\ln R^{(i)}$ showed roughly linear relations (see Figs. 1–9). For each of the plots, we estimated a simple regression model by ordinary least squares. Although the assumptions of the regression model are clearly not met, the plots show that the regression lines do seem to give an adequate summary of the relation between $\ln(S^{(i)} - R^{(i)})$ versus $\ln R^{(i)}$.

Table 1
Exact upper percentage points for various random variables $S = T(B)$, where B is the Brownian bridge

T	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.01$
T_{Kol}	1.225	1.359	1.632
T_{CvM}	0.5893	0.6792	0.8622
T_{AD}	1.3903	1.5786	1.9621

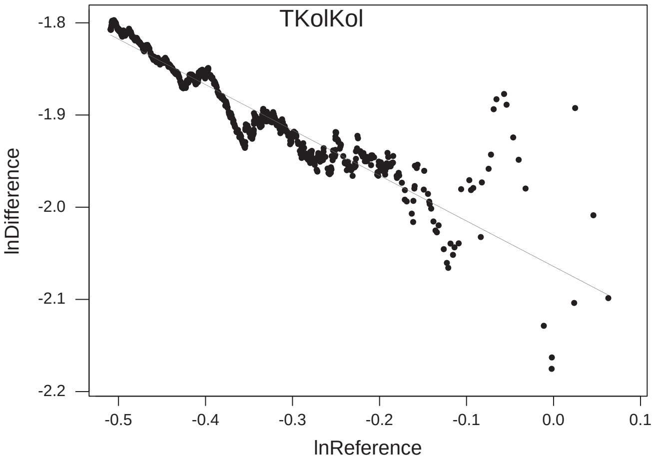

Fig. 1. Plot of $\ln(S^{(i)} - R^{(i)})$ versus $\ln R^{(i)}$ for $i = 9001, \ldots, 10\,000$, where $S^{(i)}$ is the ith order statistic of $S = T_{\mathrm{Kol}} \circ T_{\mathrm{Kol}}(\Gamma)$ and $R^{(i)}$ is the ith order statistic of $R = ||T_{\mathrm{Kol}}||_{\mathscr{H}_B} \cdot T_{\mathrm{Kol}}(B_{\mathrm{Kol}})$, based on a random sample of length 10 000 taken from the distribution of Γ. The least squares line indicates that the critical value of the size α test based on T may be approximated by $0.5000(t^{\alpha}_{\mathrm{Kol}}) + 0.17864(t^{\alpha}_{\mathrm{Kol}})^{-0.4932}$.

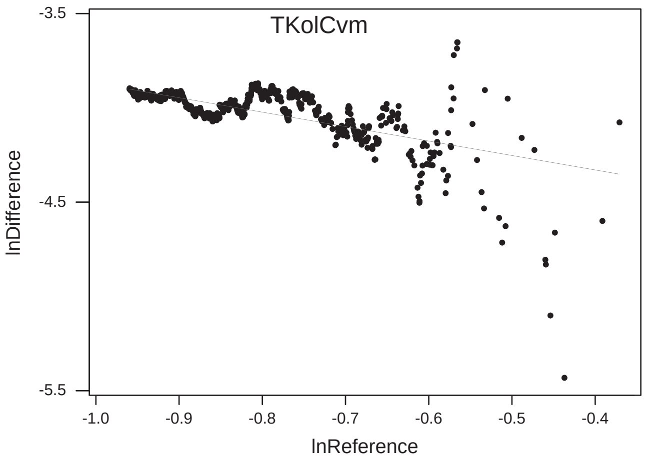

Fig. 2. Plot of $\ln(S^{(i)} - R^{(i)})$ versus $\ln R^{(i)}$ for $i = 9001, \ldots, 10\,000$, where $S^{(i)}$ is the ith order statistic of $S = T_{\mathrm{Kol}} \circ T_{\mathrm{CvM}}(\Gamma)$ and $R^{(i)}$ is the ith order statistic of $R = ||T_{\mathrm{CvM}}||_{\mathscr{H}_B} \cdot T_{\mathrm{Kol}}(B_{\mathrm{CvM}})$, based on a random sample of length 10 000 taken from the joint distribution of Γ. The least squares line indicates that the critical value of the size α test based on T may be approximated by $0.3183(t^{\alpha}_{\mathrm{Kol}}) + 0.02331(t^{\alpha}_{\mathrm{Kol}})^{-0.7660}$.

From this relation we may deduce that there exist constants c and p such that

$$S^{(i)} \approx R^{(i)} + c(R^{(i)})^{-p},$$

and infer for $0 < \alpha \leqslant 10\%$ that a similar approximation holds between the α upper percentage points of S and $T_2(B_1)$. Table 2 summarizes the approximations found in

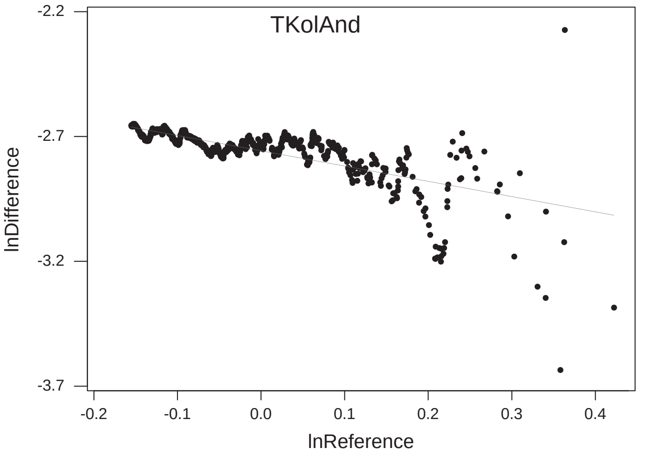

Fig. 3. Plot of $\ln(S^{(i)} - R^{(i)})$ versus $\ln R^{(i)}$ for $i = 9001, \ldots, 10\,000$, where $S^{(i)}$ is the ith order statistic of $S = T_{\mathrm{Kol}} \circ T_{\mathrm{AD}}(\Gamma)$ and $R^{(i)}$ is the ith order statistic of $R = \|T_{\mathrm{AD}}\|_{\mathscr{H}_B} \cdot T_{\mathrm{Kol}}(B_{\mathrm{AD}})$, based on a random sample of length 10 000 taken from the joint distribution of Γ. The least squares line indicates that the critical value of the size α test based on T may be approximated by $0.7071(t^{\alpha}_{\mathrm{Kol}}) + 0.07845(t^{\alpha}_{\mathrm{Kol}})^{-0.6116}$.

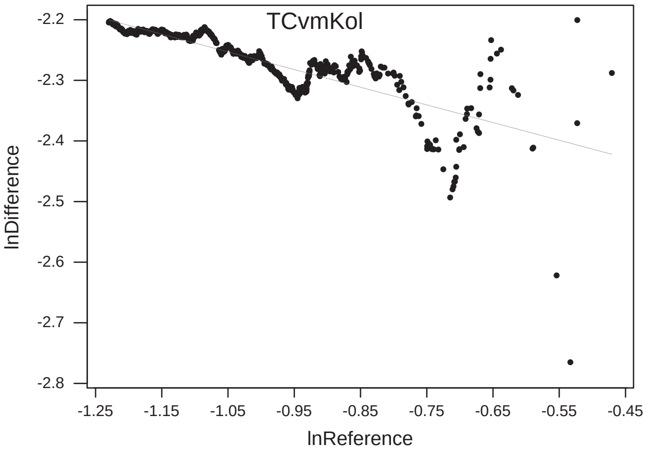

Fig. 4. Plot of $\ln(S^{(i)} - R^{(i)})$ versus $\ln R^{(i)}$ for $i = 9001, \ldots, 10\,000$, where $S^{(i)}$ is the ith order statistic of $S = T_{\mathrm{CvM}} \circ T_{\mathrm{Kol}}(\Gamma)$ and $R^{(i)}$ is the ith order statistic of $R = \|T_{\mathrm{Kol}}\|_{\mathscr{H}_B} \cdot T_{\mathrm{CvM}}(B_{\mathrm{Kol}})$, based on a random sample of length 10 000 taken from the joint distribution of Γ. The least squares line indicates that the critical value of the size α test based on T may be approximated by $0.5000(t^{\alpha}_{\mathrm{CvM}}) + 0.09469(t^{\alpha}_{\mathrm{CvM}})^{-0.2917}$.

Figs. 1–9, and evaluates the approximations for $\alpha = 0.10$, $\alpha = 0.05$ and $\alpha = 0.01$. Observe that the approximated 0.10, 0.05 and 0.01 upper percentage points given for $T_{\mathrm{Kol}} \circ T_{\mathrm{CvM}}(\Gamma)$ are quite close to those given for $T_{\mathrm{CvM}} \circ T_{\mathrm{Kol}}(\Gamma)$. The same holds for $T_{\mathrm{Kol}} \circ T_{\mathrm{AD}}(\Gamma)$ and $T_{\mathrm{AD}} \circ T_{\mathrm{Kol}}(\Gamma)$, and for $T_{\mathrm{CvM}} \circ T_{\mathrm{AD}}(\Gamma)$ and $T_{\mathrm{AD}} \circ T_{\mathrm{CvM}}(\Gamma)$.

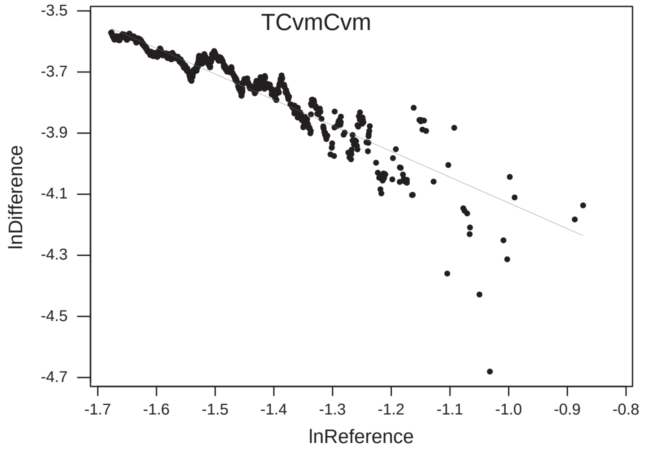

Fig. 5. Plot of $\ln(S^{(i)} - R^{(i)})$ versus $\ln R^{(i)}$ for $i = 9001, \ldots, 10\,000$, where $S^{(i)}$ is the ith order statistic of $S = T_{\mathrm{CvM}} \circ T_{\mathrm{CvM}}(\Gamma)$ and $R^{(i)}$ is the ith order statistic of $R = ||T_{\mathrm{CvM}}||_{\mathscr{H}_B} \cdot T_{\mathrm{CvM}}(B_{\mathrm{CvM}})$, based on a random sample of length 10 000 taken from the distribution of Γ. The least squares line indicates that the critical value of the size α test based on T may be approximated by $0.3183(t^{\alpha}_{\mathrm{CvM}}) + 0.01820(t^{\alpha}_{\mathrm{CvM}})^{-0.8440}$.

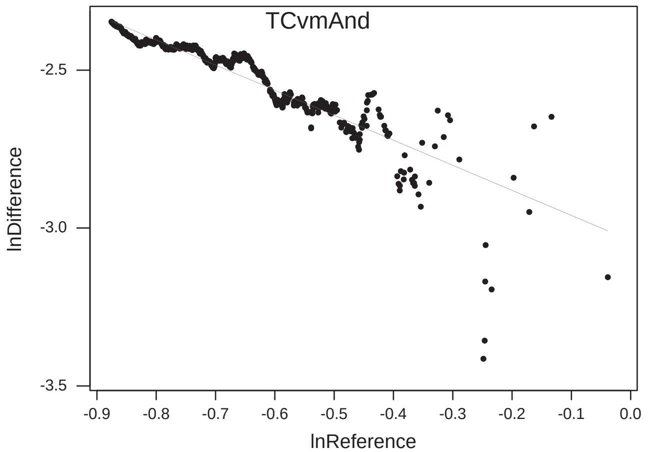

Fig. 6. Plot of $\ln(S^{(i)} - R^{(i)})$ versus $\ln R^{(i)}$ for $i = 9001, \ldots, 10\,000$, where $S^{(i)}$ is the ith order statistic of $S = T_{\mathrm{CvM}} \circ T_{\mathrm{AD}}(\Gamma)$ and $R^{(i)}$ is the ith order statistic of $R = ||T_{\mathrm{AD}}||_{\mathscr{H}_B} \cdot T_{\mathrm{CvM}}(B_{\mathrm{AD}})$, based on a random sample of length 10 000 taken from the distribution of Γ. The least squares line indicates that the critical value of the size α test based on T may be approximated by $0.7071(t^{\alpha}_{\mathrm{CvM}}) + 0.06303(t^{\alpha}_{\mathrm{CvM}})^{-0.7935}$.

Table 2 seems to suggest that the rate at which $y^{-2} \log P(S > y)$ converges to a constant $-a/2$ (recall (2)) is relatively slow for functionals T involving T_{Kol}.

The random variable $T_{\mathrm{CvM}} \circ T_{\mathrm{CvM}}(\Gamma)$ is the only one occurring in Table 2 which has been tabulated [6] see also [9]. For this random variable the exact 0.10,0.05 and

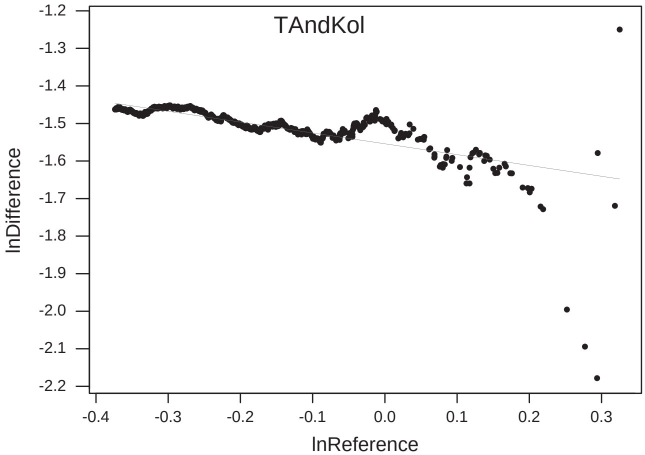

Fig. 7. Plot of $\ln(S^{(i)} - R^{(i)})$ versus $\ln R^{(i)}$ for $i = 9001, \ldots, 10\,000$, where $S^{(i)}$ is the ith order statistic of $S = T_{\mathrm{AD}} \circ T_{\mathrm{Kol}}(\Gamma)$ and $R^{(i)}$ is the ith order statistic of $R = ||T_{\mathrm{Kol}}||_{\mathscr{H}_B} \cdot T_{\mathrm{AD}}(B_{\mathrm{Kol}})$, based on a random sample of length 10 000 taken from the distribution of Γ. The least squares line indicates that the critical value of the size α test based on T may be approximated by $0.5000(t^{\alpha}_{\mathrm{AD}}) + 0.25823(t^{\alpha}_{\mathrm{AD}})^{-0.2887}$.

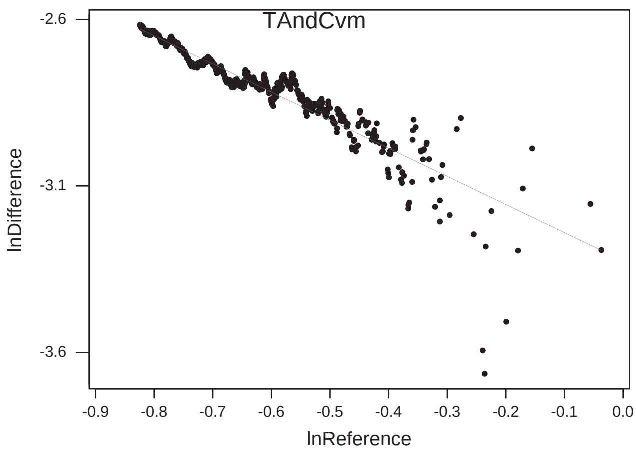

Fig. 8. Plot of $\ln(S^{(i)} - R^{(i)})$ versus $\ln R^{(i)}$ for $i = 9001, \ldots, 10\,000$, where $S^{(i)}$ is the ith order statistic of $S = T_{\mathrm{AD}} \circ T_{\mathrm{CvM}}(\Gamma)$ and $R^{(i)}$ is the ith order statistic of $R = ||T_{\mathrm{CvM}}||_{\mathscr{H}_B} \cdot T_{\mathrm{AD}}(B_{\mathrm{CvM}})$, based on a random sample of length 10 000 taken from the distribution of Γ. The least squares line indicates that the critical value of the size α test based on T may be approximated by $0.3183(t^{\alpha}_{\mathrm{AD}}) + 0.09464(t^{\alpha}_{\mathrm{AD}})^{-0.8447}$.

0.01 upper percentage points, respectively, are $\sqrt{0.04694} = 0.2167$, $\sqrt{0.05840} = 0.2417$ and $0.08685 = 0.2947$, so the approximation given in Table 2 seems to be quite accurate.

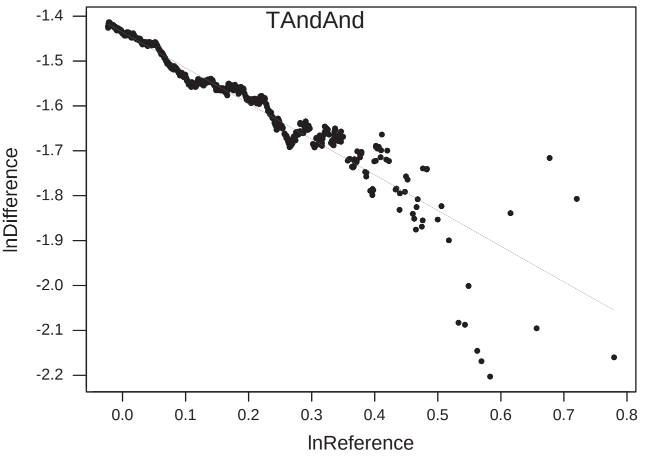

Fig. 9. Plot of $\ln(S^{(i)} - R^{(i)})$ versus $\ln R^{(i)}$ for $i = 9001, \ldots, 10,000$, where $S^{(i)}$ is the ith order statistic of $S = T_{AD} \circ T_{AD}(\Gamma)$ and $R^{(i)}$ is the ith order statistic of $R = ||T_{AD}||_{\mathscr{H}B}$. $T_{AD}(B_{AD})$, based on a random sample of length 10000 taken from the distribution of Γ. The least squares line indicates that the critical value of the size α test based on T may be approximated by $0.7071(t^{\alpha}_{AD}) + 0.31340(t^{\alpha}_{AD})^{-0.7945}$.

Table 2
Approximation of upper percentage points for various random variables $S = T(\Gamma)$, where Γ is the Brownian pillow

T	$0<\alpha\leqslant 0.10$	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.01$
$T_{Kol} \circ T_{Kol}$	$0.5000(t^{\alpha}_{Kol}) + 0.17864(t^{\alpha}_{Kol})^{-0.4932}$	0.7741	0.8331	0.9563
$T_{Kol} \circ T_{CvM}$	$0.3183(t^{\alpha}_{Kol}) + 0.02331(t^{\alpha}_{Kol})^{-0.7660}$	0.4099	0.4510	0.5355
$T_{Kol} \circ T_{AD}$	$0.7071(t^{\alpha}_{Kol}) + 0.07845(t^{\alpha}_{Kol})^{-0.6616}$	0.9355	1.0260	1.2121
$T_{CvM} \circ T_{Kol}$	$0.5000(t^{\alpha}_{CvM}) + 0.09469(t^{\alpha}_{CvM})^{-0.2917}$	0.4051	0.4456	0.5300
$T_{CvM} \circ T_{CvM}$	$0.3183(t^{\alpha}_{CvM}) + 0.01820(t^{\alpha}_{CvM})^{-0.8440}$	0.2160	0.2414	0.2951
$T_{CvM} \circ T_{AD}$	$0.7071(t^{\alpha}_{CvM}) + 0.06303(t^{\alpha}_{CvM})^{-0.7935}$	0.5126	0.5660	0.6806
$T_{AD} \circ T_{Kol}$	$0.5000(t^{\alpha}_{AD}) + 0.25823(t^{\alpha}_{AD})^{-0.2887}$	0.9300	1.0157	1.1936
$T_{AD} \circ T_{CvM}$	$0.3183(t^{\alpha}_{AD}) + 0.09464(t^{\alpha}_{AD})^{-0.8447}$	0.5142	0.5668	0.6781
$T_{AD} \circ T_{AD}$	$0.7071(t^{\alpha}_{AD}) + 0.31340(t^{\alpha}_{AD})^{-0.7945}$	1.2243	1.3343	1.5709

In [9] the use of Cornish-Fisher expansions to approximate upper percentage points of $T_{CvM} \circ T_{CvM}(\Gamma)$ is advocated. However, Cornish–Fisher expansions typically yield inaccurate results for α tending to zero. Recall that the situation where α tends to zero is of considerable theoretical interest.

5. Possible generalizations

In this section we address the question whether it is possible to generalize the key result Proposition 3.2.1 for a wider class of functionals.

Let T_1, T_2 be positive homogeneous bounded functionals defined on Hilbert spaces $\mathscr{H}_1, \mathscr{H}_2$, respectively. Let T be a functional defined on the tensor product $\mathscr{H}_1 \circ \mathscr{H}_2$ which satisfies $T(f_1 \circ f_2) = T_1(f_1)T_2(f_2)$ for all $f_1 \in \mathscr{H}_1$ and $f_2 \in \mathscr{H}_2$. Then T is said to possess the product property if $||T|| = ||T_1|| \cdot ||T_2||$.

First note that the product property entirely depends on the way we extend the functional $T_1 \circ T_2$ from basic elements $\phi_{1j} \circ \phi_{2k}$, where it is equal to $T_1(\phi_{1j})T_2(\phi_{2k})$, onto the whole tensor product $\mathscr{H}_1 \circ \mathscr{H}_2$. As we saw in Proposition 3.2.1 the product property holds if the functionals are extended bilinearly.

If both T_i are of the form $T_i(f) = |G(f, \dots, f)|^{1/k}$ for some symmetric k-linear forms T_1, T_2, then the tensor product $T_1 \circ T_2$ are naturally defined by k-linearity. For $k = 1, 2$, the k-linear extension respects the product property (for $k = 2$ this follows from Proposition 3.2.1, for $k = 1$ this also does, because the modulus of a linear functional is the square root of its square, which is a bilinear form). However, for $k \geqslant 3$ this no longer holds, as the next example shows.

Example 7 (Trilinear forms). Consider the trilinear form $G(f_1, f_2, f_3)$ defined on the space R^2 as follows: $G(e_1, e_1, e_2) = G(e_1, e_2, e_1) = G(e_2, e_1, e_1) = 1, G(e_2, e_2, e_2) = -6/5$, and $G(e_{i_1}, e_{i_2}, e_{i_3}) = 0$ for all other combination of indices (here e_1, e_2 are basic vectors). Thus for a vector $f = (x, y)$ we have $G(f, f, f) = 3x^2y - \frac{6}{5}y^3$. Put $T_1(f) = T_2(f) = |G(f, f, f)|^{1/3}$. It follows that $||T_1|| = ||T_2|| = T_1(0, -1) = (6/5)^{\frac{1}{3}}$. However, since $||T_1 \circ T_2|| = T_1 \circ T_2(\frac{\sqrt{14}}{\sqrt{39}}, 0, 0, \frac{5}{\sqrt{39}}) = (\frac{10}{\sqrt{39}})^{\frac{1}{3}}$, it also follows that $||T_1 \circ T_2|| = (10/\sqrt{39})^{\frac{1}{3}} > (36/25)^{\frac{1}{3}} = ||T_1|| \cdot ||T_2||$.

Some other extensions may violate the product property even in the simplest cases.

Example 8 (Other extensions). Let $d \geqslant 2$, and consider two functions T_1, T_2 defined on $\mathbb{R}^d$ by $T_1(f) = T_2(f) = ||f||$ for $f \in \mathbb{R}^d$. Let us show that there exists a positively homogeneous sublinear functional T defined on $\mathbb{R}^{d^2}$ such that $T(f_1 \circ f_2) = T(f_1)T_2(f_2)$ for every $f_1, f_2 \in \mathbb{R}^d$, but $||T|| > ||T_1|| \cdot ||T_2||$. Let $\mathscr{S}_{d-1}$ denote the unit sphere in $\mathbb{R}^d$, and consider

$$\mathscr{S}_{d-1} \circ \mathscr{S}_{d-1} = \{f \circ g = fg^T \colon f, g \in \mathscr{S}_{d-1}\}.$$

Obviously, $\mathscr{S}_{d-1} \circ \mathscr{S}_{d-1}$ is a compact subset of $\mathscr{S}_{d^2-1}$, which does not coincide with $\mathscr{S}_{d^2-1}$ (any $f \in \mathscr{S}_{d-1} \circ \mathscr{S}_{d-1}$, considered as $d \times d$-matrix, has rank 1). Take any $g \in \mathscr{S}_{d^2-1} \backslash \mathscr{S}_{d-1} \circ \mathscr{S}_{d-1}$. From compactness it follows that $\inf_{f \in \mathscr{S}_{d-1} \circ \mathscr{S}_{d-1}} ||g - f|| = \alpha > 0$. Therefore g is not an element of $\mathrm{conv}(\mathscr{S}_{d-1} \circ \mathscr{S}_{d-1})$, and hence the convexity and compactness of $\mathrm{conv}(\mathscr{S}_{d-1} \circ \mathscr{S}_{d-1})$ implies $T(g) > 1$, where T is the Minkowski functional defined by

$$T(g) = (\sup\{\lambda \in \mathbb{R}_+ \colon \lambda g \in \operatorname{conv}(\mathscr{S}_{d-1} \circ \mathscr{S}_{d-1})\})^{-1}.$$

It is clear that T is sublinear and positive homogeneous. We have $T(f_1 \circ f_2) = 1$ for all $f_1, f_2 \in S_{d-1}$, hence by sublinearity $T(f_1 \circ f_2) = ||f_1|| \cdot ||f_2||$ for all $f_1, f_2 \in \mathbb{R}^d$. Taking $T_1 = T_2 = || \cdot ||_{\mathbb{R}^d}$, we see that $||T|| \geqslant |T(g)| > 1 = ||T_1|| \cdot ||T_2||$.

On the other hand, we have the following positive result: for any pair of sublinear positively homogeneous functionals T_1, T_2 defined on Hilbert spaces $\mathscr{H}_1$, $\mathscr{H}_2$ there does exist an extension (of its tensor product) that respects the product property. Indeed, using sublinearity we get $T_i(f) = \sup_{f_i^* \in B_i^*} |L_{f_i^*}(f)|$, where $L_{f_i^*}$ denotes the linear form $\langle f, f_i^* \rangle$, and B_i^* is a polar to the set $B_i = \{f_i \colon |T_i(f_i)| \leqslant 1\}$; the polar B_i^* of a subset $B_i \subset \mathscr{H}_i$ is defined as $B_i^* = \{f_i^* \in \mathscr{H}_i, \sup_{f \in B_i} \langle f, f_i^* \rangle \leqslant 1\}$ (cf. [8]). We can now define the extension $T_1 \circ T_2$ by

$$T_1 \circ T_2 = \sup_{f_1^* \in B_1^*} \sup_{f_2^* \in B_2^*} |L_{f_1^*} \circ L_{f_2^*}|, \tag{15}$$

where $L_{f_1^*} \circ L_{f_2^*}$ is the uniquely defined linear form satisfying

$$L_{f_1^*} \circ L_{f_2^*}(f_1 \circ f_2) = L_{f_1^*}(f_1) \cdot L_{f_2^*}(f_2)$$

for all $f_1 \in \mathscr{H}_1$ and $f_2 \in \mathscr{H}_2$. Now, it is easy to verify that $||T_1 \circ T_2||_{\mathscr{H}_1 \circ \mathscr{H}_2} = ||T_1||_{\mathscr{H}_1} ||T_2||_{\mathscr{H}_2}$, which actually follows from Corollary 3.1 by setting $V_i = B_i^*$, $v_i = f_i^*$, and $Q_{v_i}(f, g) = L_{f_i^*}(f) \cdot L_{f_i^*}(g)$. Thus, for any pair of functionals T_1, T_2 extension (15) respects the product property. However, this extension is not always natural and suitable. For instance, it does not coincide with the bilinear extension when both T_1, T_2 are square roots of positively semidefinite bilinear forms.

6. Proofs

Proof of Lemma 2. Statement (i) is well known, we include the proof for convenience of the reader. Consider $v \in V$ and $f \in \mathscr{H}$. Since the normalized eigenfunctions of $\mathscr{A}_v$ form a complete orthonormal basis in $\mathscr{H}$, we may write

$$f = \sum_{j=1}^{\infty} b_j \phi_{j,v}, \quad \mathscr{A}_v f = \sum_{j=1}^{\infty} b_j \lambda_{j,v} \phi_{j,v},$$

and hence

$$\langle f, \mathscr{A}_v f \rangle_{\mathscr{H}} = \sum_{j=1}^{\infty} (b_j)^2 \lambda_{j,v} \leqslant \lambda_{1,v} \sum_{j=1}^{\infty} (b_j)^2. \tag{16}$$

Thus,

$$\sup_{f \in \mathcal{O}_{\mathscr{H}}} Q_v(f, f) = \lambda_{1,v} = Q_v(\phi_{1,v}, \phi_{1,v}),$$

where $\mathcal{O}_{\mathscr{H}}$ is the unit ball in $\mathscr{H}$. This yields

$$||T||^2_{\mathscr{H}} = \sup_{f \in \mathcal{O}_{\mathscr{H}}} \sup_{v \in V} Q_v(f,f) = \sup_{v \in V} \lambda_{1,v} = \lambda_{1,w}.$$

This completes the proof of Lemma 2(i).

Next, we turn to statement (ii). According to Theorem 3.7 in [1], we have the "principal components decomposition"

$$X = \sum_{j=1}^{\infty} Z_j \phi_{j,w}$$

with probability 1, where $Z_1, Z_2, \ldots$ is a sequence of independent standard normal random variables (principal components decomposition was first applied to Gaussian processes in [25]). Remark that the random variables $Z_1, Z_2, \ldots$ may be retrieved from X by

$$Z_j = \langle X, \phi_{j,w} \rangle_{\mathscr{H}},$$

and observe that Z coincides with Z_1. It follows from (16) that

$$\langle X, \mathscr{A}_v X \rangle_{\mathscr{H}} = \sum_{j=1}^{\infty} Z_j^2 \lambda_{j,w} \geqslant Z^2 \lambda_{1,w}$$

with probability 1, and hence Lemma 2(i) yields

$$T(X) \geqslant \sqrt{Q_w(X)} \geqslant |Z| \sqrt{\lambda_{1,w}} = |Z| \cdot T(\phi_{1,w})$$

with probability 1. This concludes the proof of Lemma 2(ii). □

Proof of Lemma 3. First, we verify that the linear space with the introduced scalar product is indeed a Hilbert space. Then by a direct calculation we show that

$$\langle f, K_B(\cdot, t) \rangle = f(t) \quad \text{for all } f \in \mathscr{H}_B.$$

Thus, the reproducing property with kernel $K_B(s,t) = \min(s,t) - st$ holds in this space. This concludes the proof. □

Proof of Lemma 4. For $i = 1, 2$, choose an arbitrary orthonormal base $\{h_{ik}\}$ in $\mathscr{H}_i$. Consider the Hilbert space $\mathscr{H}$ that consists of all functions $f(s,t) = \sum_{j,k=1}^{\infty} a_{jk} \cdot h_{1j}(s) h_{2k}(t)$ with $\sum_{j,k=1}^{\infty} a_{jk}^2 < \infty$ and equipped with the scalar product

$$\left\langle \sum_{j_1,k_1} a_{j_1,k_1} \cdot h_{1j} h_{2k}, \sum_{j_2,k_2} b_{j_2 k_2} \cdot h_{1j} h_{2k} \right\rangle = \sum_{j,k} a_{jk} b_{jk}.$$

This space is nothing else but the tensor product $\mathscr{H}_1 \circ \mathscr{H}_2$, where any element $f_1 \circ f_2$ is identified with the corresponding (usual) product $f(s,t) = f_1(s) f_2(t) \in \mathscr{H}$. Now it remains to note that all elements of the space $\mathscr{H}$ satisfy reproducing property with the kernel $K(s_1, t_1, s_2, t_2) = K_1(s_1, t_1) \cdot K_2(s_2, t_2)$. This completes the proof of Lemma 4. □

Proof of Proposition 3.2.1. Note that since $||f_i||_{\mathscr{H}_i} = 1$ implies $||f_1 \circ f_2||_{\mathscr{H}_1 \circ \mathscr{H}_2} = 1$, the inequality

$$\sup_{f \in \mathscr{H}_1 \circ \mathscr{H}_2} |Q_1 \circ Q_2(f)| \geqslant \sup_{f_1 \in \mathscr{H}_1} |Q_1(f_1)| \sup_{f_2 \in \mathscr{H}_2} |Q_2(f_2)|$$

is straightforward.

Therefore, it remains to prove the opposite inequality. To show this, assume the contrary: for some $g \in \mathscr{H}_1 \circ \mathscr{H}_2$ we have

$$|Q_1 \circ Q_2(g)| > \sup_{f_1 \in \mathscr{H}_1} |Q_1(f_1)| \sup_{f_2 \in \mathscr{H}_2} |Q_2(f_2)|.$$

Decompose g in the basis $\{h_{1i} \circ h_{2j}\}$, so $g = \sum_{i,j=1}^{\infty} \beta_{ij} h_{1i} \circ h_{2j}$, and consider the sequence $g_N = \sum_{i,j=1}^{N} \beta_{ij} h_{1i} \circ h_{2j}$, $N \geqslant 1$. Since $g_N \to g$ in the space $\mathscr{H}_1 \circ \mathscr{H}_2$, we see that

$$|Q_1 \circ Q_2(g_N)| > \sup_{f_1 \in \mathscr{H}_1} |Q_1(f_1)| \sup_{f_2 \in \mathscr{H}_2} |Q_2(f_2)| \geqslant \sup_{f_1 \in \mathscr{H}_{1N}} |Q_1(f_1)| \sup_{f_2 \in \mathscr{H}_{2N}} |Q_2(f_2)|$$

for sufficiently large N; here $\mathscr{H}_{iN}$ denotes the N-dimensional subspace of $\mathscr{H}_i$ spanned by $h_{i1}, \ldots, h_{iN}$, $i = 1, 2$. Since $g_N \in \mathscr{H}_{1N} \circ \mathscr{H}_{2N}$, it follows that

$$\sup_{f \in \mathscr{H}_{1N} \circ \mathscr{H}_{2N}} |Q_1 \circ Q_2(f)| > \sup_{f_1 \in \mathscr{H}_{1N}} |Q_1(f_1)| \sup_{f_2 \in \mathscr{H}_{2N}} |Q_2(f_2)|.$$

So, if the equality $\sup_{f \in \mathscr{H}_1 \circ \mathscr{H}_2} |Q_1 \circ Q_2(f)| = \sup_{f_1 \in \mathscr{H}_1} |Q_1(f_1)| \sup_{f_2 \in \mathscr{H}_2} |Q_2(f_2)|$ fails for infinite-dimensional spaces, then it does for suitable finite-dimensional spaces.

Thus, we assume that both $\mathscr{H}_1$ and $\mathscr{H}_2$ are of dimension N. In this case the space $\mathscr{H}_1 \circ \mathscr{H}_2$ is the space of matrices $\mathbb{R}^{N^2}$ with the scalar product $\langle F, G \rangle_{\mathscr{H}_i} = \mathrm{tr}(F^T G)$, that is the sum of diagonal elements of the matrix $F^T G$. The embedding $\{f_1 \circ f_2\} \subset \mathscr{H}_1 \circ \mathscr{H}_2$ is realized by the formula $f_1 \circ f_2 = f_1 f_2^T$. The form Q_i is given on $\mathscr{H}_i$ by the formula $Q_i(f, g) = \langle f, \mathscr{A}_i g \rangle_{\mathscr{H}_i}$, where $\mathscr{A}_i$ is a self-conjugate operator on $\mathscr{H}_i$, for which there exists a complete orthonormal system of eigenvectors $\{\phi_{ij}\}_{j=1}^{N}$ such that

$$\mathscr{A}_i \left(\sum_{j=1}^{N} a_j \phi_{ij} \right) = \sum_{j=1}^{N} \lambda_j a_j \phi_{ij},$$

where $\lambda_{i1}, \ldots, \lambda_{iN}$ are the eigenvalues of $\mathscr{A}_i$. It follows that $||Q_i||$ is equal to $||\mathscr{A}_i|| = \max_{j=1,\ldots,N} |\lambda_{ij}|$. Let $\mathscr{A}$ be the operator in $\mathscr{H}_1 \circ \mathscr{H}_2$ given by

$$\mathscr{A}(\phi_{1j} \circ \phi_{2k}) = \mathscr{A}_1 \phi_{1j} \circ \mathscr{A}_2 \phi_{2k}$$

(in the matrix representation, we have $\mathscr{A}(F) = \mathscr{A}_1 F \mathscr{A}_2^T$ for every $F \in \mathbb{R}^{N^2}$). As $\mathscr{A}(\phi_{1j} \circ \phi_{2k}) = \mathscr{A}_1 \phi_{1j} \circ \mathscr{A}_2 \phi_{2k} = \lambda_1 \phi_{1j} \circ \lambda_2 \phi_{2k}$, it follows that each $\phi_{1j} \circ \phi_{jk}$ is an eigenvector of $\mathscr{A}$ with the corresponding eigenvalue $\lambda_{1j} \lambda_{2k}$. The system $\{\phi_{1j} \circ \phi_{2k}\}_{j,k=1}^{N}$ is obviously orthonormal, and consists of N^2 vectors, and hence is a complete orthonormal system of eigenvectors of $\mathscr{A}$ in $\mathbb{R}^{N^2}$.

This implies that for the form $Q_1 \circ Q_2$ defined by

$$Q_1 \circ Q_2(x, y) = \langle x, \mathscr{A} y \rangle_{\mathscr{H}_1 \circ \mathscr{H}_2},$$

we have

$$\begin{aligned}||Q_1 \circ Q_2|| = ||\mathscr{A}|| &= \max_{j,k=1,\dots,N} |\lambda_{1j}\lambda_{2k}| = \max_{j=1,\dots,N} |\lambda_{1j}| \cdot \max_{k=1,\dots,N} |\lambda_{2k}| \\ &= ||\mathscr{A}_1|| \cdot ||\mathscr{A}_2|| = ||Q_1|| \cdot ||Q_2||.\end{aligned}$$

Thus, in the finite-dimensional case the statement holds, which concludes the proof of Proposition 3.2.1. □

Remark 1. Passing to the finite-dimensional case was essential in the proof of Proposition 3.2.1, because the infinite-dimensional operator $\mathscr{A}_i$ defining the form $Q_i(f, g) = \langle f, \mathscr{A} g \rangle_{\mathscr{H}_i}$ on $\mathscr{H}_i$ may not have a complete system of eigenvectors.

Proof of Lemma 5. According to Theorem 3.7 in [1], we have

$$X = \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} Z_{jk} \phi_{1j,w} \phi_{2k,w}$$

with probability 1, where the Z_{jk}'s are independent standard normal random variables. It follows that

$$\langle X(\cdot, t_2), \phi_{11,w} \rangle_{\mathscr{H}_1} = \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} Z_{jk} \phi_{2k,w}(t_2) \cdot \langle \phi_{1j,w}, \phi_{11,w} \rangle_{\mathscr{H}_1} = \sum_{k=1}^{\infty} Z_{1k} \phi_{2k,w}(t_2)$$

for every $t_2 \in M_2$, with probability 1. Observe that the RHS of the latter equation is an expansion of a mean zero Gaussian process with covariance function $K_2(s_2, t_2)$. This concludes the proof of Lemma 5. □

Proof of Lemma 6. The proof is realized in the same way as one of Lemma 3. □

References

[1] R.J. Adler, An Introduction to Continuity, Extrema, and Related Topics for General Gaussian Processes, Institute of Mathematical Statistics, Hayward, CA, 1990.

[2] T.W. Anderson, D.A. Darling, Asymptotic theory of certain "goodness of fit" criteria based on stochastic processes, Ann. Math. Statist. 23 (1952) 193–212.

[3] T.W. Anderson, D.A. Darling, A test of goodness of fit, J. Amer. Statist. Assoc. 49 (1954) 765–769.

[4] N. Aronszajn, Theory of reproducing kernels, Trans. Amer. Math. Soc. 68 (1950) 337–404.

[5] R.R. Bahadur, Stochastic comparison of tests, Ann. Math. Statist. 31 (1960) 276–295.

[6] J.R. Blum, J. Kiefer, M. Rosenblatt, Distribution free tests of independence based on the sample distribution function, Ann. Math. Statist. 32 (1961) 485–498.

[7] C. Borell, The Brunn-Minkowski inequality in Gauss space, Invent. Math. 30 (1975) 207–216.

[8] J.M. Borwein, A.S. Lewis, Convex Analysis and Nonlinear Optimization, Theory and Examples, Springer, New York, 2000.

[9] D.S. Cotterill, M. Csörgő, On the limiting distribution of and critical values for the Hoeffding, Blum, Kiefer, Rosenblatt independence criterion, Statist. Decisions 3 (1985) 1–48.

[10] M. Csörgő, Strong approximations of the Hoeffding, Blum, Kiefer, Rosenblatt multivariate empirical process, J. Multivariate Anal. 9 (1979) 84–100.
[11] M. Csörgő, L. Horváth, Limit Theorems in Change-Point Analysis, Wiley, New York, 1997.
[12] P. Deheuvels, An asymptotic decomposition for multivariate distribution-free tests of independence, J. Multivariate Anal. 11 (1981) 102–113.
[13] T. De Wet, Cramér–von Mises tests for independence, J. Multivariate Anal. 10 (1980) 38–50.
[14] D. Dugué, Sur des tests d'indépendence "indépendants de la loi", C. R. Séances Acad. Sci. Ser. A 281 (1975) 1103–1104.
[15] J. Durbin, Distribution Theory for Tests Based on the Sample Distribution Function, SIAM, Philadelphia, PA, 1973.
[16] J. Durbin, M. Knott, Components of Cramér–von Mises statistics, I, J. Roy. Statist. Soc. B 34 (1972) 290–307.
[17] B.J. Eastwood, V.R. Eastwood, Percentiles for Cramér–von Mises functionals of Gaussian processes and some applications to Bayesian tests, Stochastic Process. Appl. 42 (1992) 329–344.
[18] J.H.J. Einmahl, I. van Keilegom, Goodness-of-fit tests in nonparametric regression, 2003, in preparation.
[19] G.G. Gregory, On efficiency and optimality of quadratic tests, Ann. Statist. 8 (1980) 116–131.
[20] G. Helmberg, Introduction to Spectral Theory in Hilbert Space, North-Holland, Amsterdam, 1969.
[21] N.L. Hjort, A.J. Koning, Constancy of distributions: nonparametric monitoring of probability distributions over time, 2001, Report 2001-50, Econometric Institute.
[22] W. Hoeffding, A nonparametric test of independence, Ann. Math. Statist. 19 (1948) 546–557.
[23] T. Inglot, T. Ledwina, On probabilities of excessive deviations for Kolmogorov-Smirnov, Crámer–von Mises and chi-square statistics, Ann. Statist. 18 (1990) 1491–1495.
[24] T. Inglot, T. Ledwina, Moderately large deviations and expansions of large deviations for some functionals of weighted empirical process, Ann. Probab. 21 (1993) 1691–1705.
[25] M. Kac, A.J.F. Siegert, An explicit representation of a stationary Gaussian process, Ann. Math. Statist. 18 (1947) 438–442.
[26] W.C.M. Kallenberg, Intermediate efficiency, theory and examples, Ann. Statist. 11 (1983) 170–182.
[27] W.C.M. Kallenberg, A.J. Koning, On Wieand's theorem, Statist. Probab. Lett. 25 (1995) 121–132.
[28] L.V. Kantorovich, G.P. Akilov, Functional Analysis, Pergamon Press, Oxford-Elmsford, NY, 1982.
[29] J. Kiefer, J. Wolfowitz, On the deviations of the empiric distribution function of vector chance variables, Trans. Amer. Math. Soc. 87 (1958) 173–186.
[30] A.A. Kirillov, A.D. Gvishiani, Theorems and Problems in Functional Analysis, Springer, New York, Berlin, 1982.
[31] A.N. Kolmogorov, Sulla determinazione empirica di una legge di distribuzione, Giorn. Ist. Ital. Attuari. 4 (1933) 83–91.
[32] A.N. Kolmogorov, S.V. Fomīn, Introductory Real Analysis, Dover Publications Inc., New York, 1975.
[33] J. Komlós, P. Major, G. Tusnády, An approximation of partial sums of independent RV's and sample D.F. I, Z. Wahrsch. Verwante Gebiete 32 (1975) 111–131.
[34] A.J. Koning, Approximation of stochastic integrals with applications to goodness-of-fit tests, Ann. Statist. 20 (1992) 428–454.
[35] A.J. Koning, Approximation of the basic martingale, Ann. Statist. 22 (1994) 565–579.
[36] Ya. Nikitin, Asymptotic Efficiency of Nonparametric Tests, Cambridge University Press, Cambridge, 1995.
[37] V.I. Piterbarg, Asymptotic Methods in the Theory of Gaussian Processes and Fields, American Mathematical Society, Providence, RI, 1996.
[38] H. Rubin, J. Sethuraman, Bayes risk efficiency, Sankhyā Ser. A 27 (1965) 347–356.
[39] G. Samorodnitsky, Probability tails of Gaussian extrema, Stochastic Process. Appl. 38 (1991) 55–84.
[40] R.J. Serfling, Approximation Theorems of Mathematical Statistics, Wiley, New York, 1980.
[41] G.R. Shorack, J.A. Wellner, Empirical Processes with Applications to Statistics, Wiley, New York, 1986.

[42] A.W. van der Vaart, J.A. Wellner, Weak Convergence and Empirical Processes, with Applications to Statistics, Springer, New York, 1996.
[43] E.T. Whittaker, G.N. Watson, A Course of Modern Analysis, Cambridge University Press, Cambridge, 1927.
[44] H.S. Wieand, A condition under which the Pitman and Bahadur approaches to efficiency coincide, Ann. Statist. 4 (1976) 1003–1011.