Statistica Sinica **20** (2010), 1333-1361

RIGID MOTION INVARIANT TWO-SAMPLE TESTS

L. Baringhaus and C. Franz

Leibniz Universität Hannover

Abstract: New rigid motion invariant tests for the multivariate two-sample problem are proposed. The test statistic is based on the inter-point distances between the two samples and the inter-point distances within each sample. The asymptotic null distribution of the test statistic is a weighted sum of squares of independent unit normal random variables, the weights being the eigenvalues of a certain Hilbert-Schmidt-operator depending on the unknown underlying distribution. An estimate of the limit distribution is obtained by replacing the unknown weights by the eigenvalues of a bootstrapped version of the operator. Quantiles of the estimate are chosen as critical values. The tests are shown to be consistent. Approximate Bahadur efficiencies computed for normal location alternatives, normal scale alternatives, and Lehmann's contaminated alternative are seen to coincide locally with Pitman efficiencies. The results are supported by a simulation study.

Key words and phrases: Bootstrap in the limit, Cramér test, multivariate two-sample tests, rigid motion invariance.

1. Introduction

There is a vast literature dealing with tests for more or less special two-sample problems. Here we concentrate on multivariate nonparametric two-sample problems. To be specific, let $\mathcal{F}$ be some nonparametric family of distributions on the Borel sets $\mathcal{B}^d$ of $\mathbb{R}^d$. Let $X_1, \ldots, X_m, Y_1, \ldots, Y_n$ be independent random (column) d-vectors. Let the $X_1, \ldots, X_m$ be identically distributed with unknown distribution $F \in \mathcal{F}$, and the $Y_1, \ldots, Y_n$ be identically distributed with unknown distribution $G \in \mathcal{F}$. For broad classes $\mathcal{F}$, we aim to develop new tests for testing

$$H : F = G,\ K : F \neq G. \tag{1.1}$$

Origin. Our starting point is the Cramér test, proposed recently by Baringhaus and Franz (2004) for the class $\mathcal{F}$ of distributions H with finite expectation $\int |x|\, dH(x)$, where $|\cdot|$ denotes the Euclidean norm on $\mathbb{R}^d$. The authors give a simple proof of the inequality

$$\begin{aligned} 0 \ \le \ & \mathrm{E}|X_1 - Y_1| - \frac{1}{2}\mathrm{E}|X_1 - X_2| - \frac{1}{2}\mathrm{E}|Y_1 - Y_2| \\ = \ & \int |x-y| \, dF \otimes G(x,y) - \frac{1}{2}\int |x-y| \, dF \otimes F(x,y) \\ & \qquad - \frac{1}{2}\int |x-y| \, dG \otimes G(x,y) \end{aligned} \tag{1.2}$$

and show that equality holds if and only if $F = G$. A test statistic that is motivated by this result is the empirical counterpart to the expression on the right side of (1.2), multiplied by $mn/(m+n)$ to get a limiting null distribution, i.e.,

$$\begin{aligned} T_{m,n} = & \frac{mn}{m+n}\Big[\int |x-y| \, dF_m \otimes G_n(x,y) - \frac{1}{2}\int |x-y| \, dF_m \otimes F_m(x,y) \\ & \qquad - \frac{1}{2}\int |x-y| \, dG_n \otimes G_n(x,y)\Big] \\ = & \frac{mn}{m+n}\Big[\frac{1}{mn}\sum_{j=1}^{m}\sum_{k=1}^{n} |X_j - Y_k| - \frac{1}{2m^2}\sum_{j=1}^{m}\sum_{k=1}^{m} |X_j - X_k| \\ & \qquad - \frac{1}{2n^2}\sum_{j=1}^{n}\sum_{k=1}^{n} |Y_j - Y_k|\Big], \end{aligned}$$

where $F_m = (1/m)\sum_{i=1}^{m}\delta_{X_i}$ and $G_n = (1/n)\sum_{j=1}^{n}\delta_{Y_j}$ (δ_a denotes the distribution degenerate at $a \in \mathbb{R}^d$) are the empirical distributions of $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$. Rejection is for large values of $T_{m,n}$, where the critical value is obtained by bootstrapping. The test statistic was proposed slightly prior to Baringhaus and Franz by Szabo et al. (2002, 2003) and also by Székely (2004). Klebanov et al. (2006) discuss the application of a permutation procedure to micro-array data analysis. A more general type of test statistic is of the form

$$\begin{aligned} T^{\ell}_{m,n} = & \frac{mn}{m+n}\Big[\int \ell(x,y) \, dF_m \otimes G_n(x,y) - \frac{1}{2}\int \ell(x,y) \, dF_m \otimes F_m(x,y) \\ & \qquad - \frac{1}{2}\int \ell(x,y \, dG_n \otimes G_n(x,y)\Big] \\ = & \frac{mn}{m+n}\Big[\frac{1}{mn}\sum_{j=1}^{m}\sum_{k=1}^{n} \ell(X_j, Y_k) - \frac{1}{2m^2}\sum_{j=1}^{m}\sum_{k=1}^{m} \ell(X_j, X_k) \\ & \qquad - \frac{1}{2n^2}\sum_{j=1}^{n}\sum_{k=1}^{n} \ell(Y_j Y_k)\Big], \end{aligned}$$

where $\ell : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is a continuous negative definite kernel. Like $T_{m,n}$, it is motivated by an inequality asserting that

$$0 \leq \mathrm{E}\ell(X_1, Y_1) - \frac{1}{2}\mathrm{E}\ell(X_1, X_2) - \frac{1}{2}\mathrm{E}\ell(Y_1, Y_2) \tag{1.3}$$

$$= \int \ell(x, y)\, dF \otimes G(x, y) - \frac{1}{2}\int \ell(x, y)\, dF \otimes F(x, y) - \frac{1}{2}\int \ell((x, y)\, dG \otimes G(x, y) \tag{1.4}$$

for distributions F, G such that ℓ is integrable with respect to $H \otimes H$ where $H = (F + G)/2$. See Berg, Christensen, and Ressel (1984, Thm. 2.1), Zinger, Klebanov, and Kakosyan (1989) and Klebanov (2005). Clearly, equality holds if $F = G$. To achieve consistency, of special interest are kernels for which the converse is true, i.e. equality holds in (1.3) if and only if $F = G$. Examples of such kernels are given in Klebanov (2005). Especially, for all $0 < r < 2$ the kernels $\ell^r(x, y) = |x - y|^r,\ x, y \in \mathbb{R}^d$, are shown to have the desired property, see e.g., Klebanov (2005) and Székely (2004).

Rigid motion invariant tests. Our interest is in tests that are rigid motion invariant. Recall that a rigid motion g on $\mathbb{R}^d$ is a map $g : \mathbb{R}^d \to \mathbb{R}^d$ of the form $g(x) = Qx + a,\ x \in \mathbb{R}^d$, where Q is an orthogonal $d \times d$-matrix, and a is a vector in $\mathbb{R}^d$. The testing problem considered by Baringhaus and Franz is invariant with respect to the group of rigid motions. The rigid motion invariance of the Cramér test and, more generally, that of the tests based on $T^{\ell^r}_{m,n}$ is obvious. There is a natural generalization. Choose some continuous function $\phi : [0, \infty) \to \mathbb{R}$ such that $\ell_\phi(x, y) = \phi(|x - y|^2), x, y \in \mathbb{R}^d$, is a negative definite kernel, and use $T^\phi_{m,n} := 2T^{\ell_\phi}_{m,n}$ where $\ell_\phi(x, y) = \phi(|x - y|^2), x, y \in \mathbb{R}^d$, is a suitable negative definite kernel as a test statistic.

Test statistic and properties. We aim to develop statistical tests that are consistent in the sense that, given any fixed alternative $F \neq G,\ F, G \in \mathcal{F}$, the power of the test tends to 1 as the sample sizes m, n tend to infinity in a suitable way. To obtain consistency of the test rejecting the hypothesis for large values of $T^\phi_{m,n}$, some conditions need to be imposed on the function ϕ chosen. We consider the continuous functions $\phi : [0, \infty) \to \mathbb{R}$ that satisfy the inequality

$$0 \ \leq\ 2\int \phi(|x - y|^2)\, dF \otimes G(x, y) - \int \phi(|x - y|^2)\, dF \otimes F(x, y) - \int \phi(|x - y|^2)\, dG \otimes G(x, y) \tag{1.5}$$

for all distributions F, G on the Borel sets of $\mathbb{R}^d$ with finite integrals $\int \phi(|x - y|^2)\, dF \otimes G(x,y)$, $\int \phi(|x-y|^2)\, dF \otimes F(x,y)$, and $\int \phi(|x-y|^2)\, dG \otimes G(x,y)$, and with equality holding if and only if $F = G$. Such functions exist, and examples will be given. We can (and do) assume without loss of generality that $\phi(0) = 0$, and that ϕ is non-negative.

The inequality (1.5) is valid for all distributions F, G with finite support if and only if ℓ_ϕ is negative definite, see e.g., Berg, Christensen, and Ressel (1984), Proposition 7.1.2. Recall that a function $f : (0,\infty) \to \mathbb{R}$ is completely monotone if $(-1)^k f^{(k)}(t) \geq 0$ for all $t > 0$ and $k = 0, 1, \ldots$. The continuous real functions ϕ on $[0,\infty)$ for which ℓ_ϕ is negative definite for each dimension d are just the functions having a completely monotone derivative $f = \phi'|_{(0,\infty)}$ on $(0,\infty)$. Essentially, this is a result of Schoenberg (1938). A more general result is given by Guo, Hu, and Sun (1993). Since equality should hold in (1.5) if and only if $F = G$, we have to exclude the functions $\phi(t) = ct,\ t \geq 0$, with some constant $c \geq 0$, as is easily checked. Thus, in what follows we are concerned with continuous functions $\phi : [0,\infty) \to [0,\infty)$ with $\phi(0) = 0$ and non-constant completely monotone derivative on $(0,\infty)$. Such functions are clearly sub-additive,

$$\phi(s+t) \leq \phi(s) + \phi(t) \quad \text{for all } s, t \geq 0,$$

which implies that

$$\phi(|x-y|^2) \leq 2\big(\phi(|x|^2) + \phi(|y|^2)\big) \quad \text{for all } x, y \in \mathbb{R}^d. \tag{1.6}$$

Introducing for given ϕ the class $\mathcal{F} = \mathcal{F}_d(\phi)$ of distributions H on $\mathcal{B}^d$ with finite integral $\int \phi(|x|^2)\, dH(x)$, we have that for all $F, G \in \mathcal{F}_d(\phi)$ the integrals in (1.5) are finite. With $\mathcal{F}_d(\phi)$ the underlying class of distributions the testing problem (1.1) is rigid motion invariant. Mattner (1990, 1997) shows that (1.5) holds true for $F, G \in \mathcal{F}_d(\phi)$. For completeness, we give a different proof in Section 2 that deals with representations of the test statistic. It has the further advantage that it yields an alternative representation of the test statistic $T^\phi_{m,n}$, useful when studying its asymptotic behavior and that of the corresponding test. The statistic $T^\phi_{m,n}$ is the empirical counterpart of the expression on the right hand side in (1.5), multiplied by $mn/(m+n)$.

The limiting null distribution of $T^\phi_{m,n}$ is derived in Section 3 by using the Central Limit Theorem for random elements in Hilbert spaces. This method of proof is different from that given by Baringhaus and Franz (2004) for the test statistic $T_{m,n}$.

The distribution of $T^\phi_{m,n}$, say $\mathcal{L}(T^\phi_{m,n}|F)$, and that of its limiting distribution, say $\mathcal{L}(T^\phi|F)$, depend on the common underlying distribution $F = G \in \mathcal{F}_d(\phi)$

in the case when the hypothesis is true. The common underlying distribution $F = G \in \mathcal{F}_d(\phi)$ is estimated by $H_{m,n} = [m/(m+n)]F_m + [n/(m+n)]G_n$, the empirical distribution of the pooled sample, which gives $\mathcal{L}(T^{\phi}_{m,n}|H_{m,n})$ as an estimator of $\mathcal{L}(T^{\phi}_{m,n}|F)$. For given $\alpha \in (0,1)$ the $(1-\alpha)$-quantile, $c_{m,n}$, of $\mathcal{L}(T^{\phi}_{m,n}|H_{m,n})$ is an estimator of the $(1-\alpha)$-quantile of $\mathcal{L}(T^{\phi}_{m,n}|F)$, and can be used as the critical value of a test. Due to computational difficulties, it may in turn be estimated by taking Monte Carlo samples from the distribution $\mathcal{L}(T^{\phi}_{m,n}|H_{m,n})$. We show in Section 4 that a 'bootstrap in the limit method' also works. By this method, $\mathcal{L}(T^{\phi}|H_{m,n})$ is used as estimator of $\mathcal{L}(T^{\phi}_{m,n}|F)$, and the $(1-\alpha)$-quantile $c_{m,n}$ of $\mathcal{L}(T^{\phi}|H_{m,n})$ is used as critical value of the test. This procedure has the advantage, that the critical values $c_{m,n}$ can be computed by numerical methods. No Monte Carlo samples are needed.

In Section 5 the performance of some new tests is investigated by means of the approximate Bahadur efficiency, which is shown to coincide locally with the Pitman efficiency. Section 6 adds a simulation study on the power of the new tests compared to that of their well-known parametric and non-parametric competitors. The empirical results obtained emphasize the theoretical findings and show that the new procedures exhibit satisfactory power. Experiences from the simulation study providing practitioners with guidance for the application of the new tests are summarized in Section 7. Lastly, Section 8 outlines an extension to the c-sample problem.

Nonparametric multivariate two-sample problems are studied by various authors. We refer to the recent paper of Baringhaus and Franz (2004) for a small overview. It is pointed out there that only a few tests share the properties of (i) being invariant with respect to rigid motions, and (ii) being consistent against any fixed alternative $F \neq G$ where F and G belong to some large nonparametric class $\mathcal{F}$ of distributions. Of special interest is the case where $\mathcal{F}$ is the class of all distributions on the Borel sets of $\mathbb{R}$. Of course, in applications dealing with data vectors of incommensurable components one would not ask for a procedure satisfying the property (i) of rigid motion invariance. Tests based on the component-wise ranks, extensively dealt with in Puri and Sen (1971), other rank like tests (see, e.g., Hettmannsperger and McKean (1998), Zuo and He (2006)) or Kolmogorov-Smirnov-type tests (Bickel (1969), Præstgaard (1995)) may be applied. Statistical tests based on the number of nearest neighbor type coincidences (Henze (1984, 1988), Schilling (1986)) are orthogonally invariant if the Euclidean distance is used, as is done in Henze (1984). Henze's test is one of the competitors the performance of which is compared with the new procedure in Section 6. It is rigid motion invariant and it is seen to be consistent against against each alternative $F \neq G$, at least if F and G are assumed to have a.e. continuous densities.

2. Representations of the Test Statistics

Let Φ be the family of continuous functions $\phi : [0,\infty) \to [0,\infty)$ with $\phi(0) = 0$ and non-constant completely monotone derivative on $(0,\infty)$. Let us first give a useful representation for the functions in Φ. A theorem of Bernstein states that a real function on the interval $(0,\infty)$ is completely monotone if and only if it is the Laplace transform of a positive Radon measure on the Borel sets of $[0,\infty)$, see, e.g., Berg, Christensen, and Ressel (1984), Corollary 4.6.14. Thus for $\phi \in \Phi$, ϕ' can be written as

$$\phi'(z) = \int_{[0,\infty)} \exp(-\lambda z)\ d\sigma_\phi(\lambda), \qquad z > 0,$$

where σ_ϕ is some positive Radon measure σ_ϕ on the Borel sets of $[0,\infty)$. It is seen that σ_ϕ is finite if and only if $\phi'(0+) < \infty$. Using the Fubini Theorem, we get

$$\phi(z) = z\sigma_\phi(\{0\}) + \int_{(0,\infty)} \frac{1}{\lambda}[1 - \exp(-\lambda z)]\ d\sigma_\phi(\lambda), \qquad z \geq 0. \tag{2.1}$$

Writing

$$\begin{aligned} d_\phi^2(F,G) := 2\int \phi(|x-y|^2)\, dF \otimes G(x,y) - \int \phi(|x-y|^2)\, dF \otimes F(x,y) \\ - \int \phi(|x-y|^2)\, dG \otimes G(x,y) \end{aligned}$$

for $F, G \in \mathcal{F}_d(\phi)$, (1.5) can be written as $d_\phi^2(F,G) \geq 0$ for each $F, G \in \mathcal{F}_d(\phi)$, and one has $T_{m,n}^\phi = [mn/(m+n)]d_\phi^2(F_m, G_n)$.

Lemma 2.1. *Let $\phi \in \Phi$, and let X, Y be random vectors with distributions $F, G \in \mathcal{F}_d(\phi)$, respectively. Then,*

$$\begin{aligned} d_\phi^2(F,G) = \int_{(0,\infty)} \int_{\mathbb{R}^d} \frac{1}{\lambda} \left|\varphi_F(\sqrt{2\lambda}z) - \varphi_G(\sqrt{2\lambda}z)\right|^2\ d\mathrm{N}_d(0, I_d)(z)\ d\sigma_\phi(\lambda) \\ + 2\sigma_\phi(\{0\})\ |\mathrm{E}X - \mathrm{E}Y|^2, \end{aligned}$$

where σ_ϕ is the positive Radon measure associated with ϕ, the functions φ_F, φ_G are the Fourier transforms of F, G, and $d\mathrm{N}_d(0, I_d)(z)$ denotes integration with respect to the d-variate normal distribution with mean vector 0 and the d-dimensional unit matrix I_d as covariance matrix. Defining the measure τ_ϕ on the Borel sets of $\mathbb{R}^d$ by

$$\tau_\phi(B) := \int_{(0,+\infty)} \int_B \frac{1}{\lambda}\ d\mathrm{N}_d(0, 2\lambda I_d)\ d\sigma_\phi(\lambda), \qquad B \in \mathcal{B}^d,$$

it is σ-finite and

$$d_\phi^2(F,G) = \int |\varphi_F(z) - \varphi_G(z)|^2 \; d\tau_\phi(z) + 2\sigma_\phi(\{0\}) \; |\mathrm{E}X - \mathrm{E}Y|^2. \tag{2.2}$$

Proof. If $\sigma_\phi(\{0\}) > 0$ it follows from (2.1) that $z^2 \le \sigma_\phi(\{0\})^{-1}\phi(|z^2|)$ for $z \ge 0$, which implies that $\mathrm{E}|X|^2$ and $\mathrm{E}|X|$ are finite. Using the representation

$$\begin{aligned}\exp(-\lambda|x-y|^2) &= \int_{\mathbb{R}^d} \exp(\mathrm{i}\sqrt{2\lambda}(x-y)'z) \; d\mathrm{N}_d(0,I_d)(z) \\ &= \int_{\mathbb{R}^d} \cos(\sqrt{2\lambda}(x-y)'z) \; d\mathrm{N}_d(0,I_d)(z), \quad x,y \in \mathbb{R}^d,\end{aligned}$$

and introducing independent random d-vectors X_1, X_2, Y_1, Y_2 with $X_1 \stackrel{\mathcal{D}}{=} X_2 \stackrel{\mathcal{D}}{=} X$ and $Y_1 \stackrel{\mathcal{D}}{=} Y_2 \stackrel{\mathcal{D}}{=} Y$, where '$\stackrel{\mathcal{D}}{=}$' means equality in distribution, we obtain

$$\begin{aligned}
&d_\phi^2(F,G) \\
&= \mathrm{E}\big[\phi(|X_1 - Y_2|^2) + \phi(|X_2 - Y_1|^2) - \phi(|X_1 - X_2|^2) - \phi(|Y_1 - Y_2|^2)\big] \\
&= \mathrm{E}\int_{(0,\infty)} \frac{1}{\lambda}\Big[-\exp(-\lambda|X_1 - Y_2|^2) - \exp(-\lambda|X_2 - Y_1|^2) \\
&\quad + \exp(-\lambda|X_1 - X_2|^2) + \exp(-\lambda|Y_1 - Y_2|^2)\Big] \; d\sigma_\phi(\lambda) \\
&\quad + \sigma_\phi(\{0\})\mathrm{E}\big[|X_1 - Y_2|^2 + |X_2 - Y_1|^2 - |X_1 - X_2|^2 - |Y_1 - Y_2|^2\big] \\
&= \mathrm{E}\int_{(0,\infty)} \Big\{\frac{1}{\lambda}\int_{\mathbb{R}^d}\Big[-\cos(\sqrt{2\lambda}(X_1 - Y_2)'z) - \cos(\sqrt{2\lambda}(X_2 - Y_1)'z) \\
&\quad + \cos(\sqrt{2\lambda}(X_1 - X_2)'z) + \cos(\sqrt{2\lambda}(Y_1 - Y_2)'z)\Big] \; d\mathrm{N}_d(0,I_d)(z)\Big\} \; d\sigma_\phi(\lambda) \\
&\quad + 2\sigma_\phi(\{0\}) \; \mathrm{E}\left[(X_1 - Y_1)'(X_2 - Y_2)\right].
\end{aligned} \tag{2.3}$$

The integrand in (2.3) is continuous and can be written as the difference of two non-negative functions,

$$\begin{aligned}&\Big[1 - \cos\Big(\sqrt{2\lambda}(X_1 - Y_2)'z\Big) + 1 - \cos\Big(\sqrt{2\lambda}(X_2 - Y_1)'z\Big)\Big] \\ &- \Big[1 - \cos\Big(\sqrt{2\lambda}(X_1 - X_2)'z\Big) + 1 - \cos\Big(\sqrt{2\lambda}(Y_1 - Y_2)'z\Big)\Big],\end{aligned}$$

the integrals of which are finite as is seen from

$$\begin{aligned}&\mathrm{E}\int_{(0,\infty)} \frac{1}{\lambda}\int_{\mathbb{R}^d}\Big[1 - \cos\Big(\sqrt{2\lambda}(X_1 - Y_2)'z\Big)\Big] \; d\mathrm{N}_d(0,I_d)(z) \, d\sigma_\phi(\lambda) \\ &\le \mathrm{E}\phi(|X_1 - Y_2|^2) < \infty,\end{aligned}$$

and corresponding inequalities for the other three terms. Thus, due to the Fubini Theorem, we can change the order of integration in (2.3) to get the representation

$$\begin{aligned} d_\phi^2(F,G) = \int_{(0,\infty)} \bigg\{ \frac{1}{\lambda} \mathrm{E} \int_{\mathbb{R}^d} & \big[\exp(\mathrm{i}\sqrt{2\lambda}X_1'z) - \exp(\mathrm{i}\sqrt{2\lambda}Y_1'z)\big] \\ & \big[\exp(-\mathrm{i}\sqrt{2\lambda}X_2'z) - \exp(-\mathrm{i}\sqrt{2\lambda}Y_2'z)\big] \; d\mathrm{N}_d(0,I_d)(z) \bigg\} \, d\sigma_\phi(\lambda) \\ & + 2\sigma_\phi(\{0\}) \; |\mathrm{E}X_1 - \mathrm{E}Y_1|^2 , \end{aligned}$$

where, due to the spherical symmetry of the $\mathrm{N}(0,I_d)$ distribution, the integrals of the additional terms in the complex-valued integrand vanish. Applying again the Fubini Theorem, and using the definition of τ_ϕ, the representation (2.2) follows. Putting $B_0 = \{0\}$, and $B_k = \{z \in \mathbb{R}^d : |z| \geq 1/k\}$, $k \in \mathbb{N}$ we have $\tau_\phi(B_0) = 0$, and

$$\tau_\phi(B_k) \leq d \; \phi(4dk^2) < +\infty \quad \text{for } k = 1, 2, \ldots,$$

which implies that τ_ϕ is σ-finite.

Remark. As is shown by Mattner (1990, 1997), d_ϕ defines a metric on $\mathcal{F}_d(\phi)$. The representation (2.2) provides a simple proof of this result. Here, we show that $d_\phi(F,G) = 0$ implies $F = G$. In fact, from $d_\phi(F,G) = 0$ it follows that $\phi_F = \phi_G$ τ_ϕ-almost everywhere. Since τ_ϕ dominates λ^d, the Lebesgue measure on $\mathcal{B}^d$, it holds that $\phi_F = \phi_G$ λ-almost everywhere. The continuity of ϕ_F and ϕ_G and the uniqueness theorem for Fourier transforms yield $F = G$.

Replacing F and G by the empirical distributions F_m and G_n based on $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$, we have

$$T_{m,n}^\phi = \frac{mn}{m+n} \left[\int |\varphi_{F_m}(z) - \varphi_{G_n}(z)|^2 \, d\tau_\phi(z) + 2\sigma_\phi(\{0\}) |\overline{X} - \overline{Y}|^2 \right],$$

where $\overline{X} = (1/m) \sum_{k=1}^m X_k$, $\overline{Y} = (1/n) \sum_{\ell=1}^n Y_\ell$, and φ_{F_m} and φ_{G_n} are the empirical Fourier transforms of $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$. Introducing for a distribution H on $\mathcal{B}^d$ the sine-cosine transform $\eta_H(z)$, $z \in \mathbb{R}^d$,

$$\eta_H(z) := \int [\cos(x'z) + \sin(x'z)] \, dH(x), \; z \in \mathbb{R}^d,$$

it is seen by arguing as in the proof of Lemma 2.1 that for distributions $F, G \in \mathcal{F}_d(\phi)$ $d_\phi^2(F,G)$ can be also written as

$$d_\phi^2(F,G) = \int |\eta_F(z) - \eta_G(z)|^2 \, d\tau_\phi(z) + 2\sigma_\phi(\{0\}) |\mathrm{E}X - \mathrm{E}Y|^2.$$

Table 1. Measures σ_ϕ for selected functions ϕ.

Function ϕ	Associated measure σ_ϕ	
$\phi(z) = 1 - \exp(-z/2)$	$\sigma_\phi = \frac{1}{2}\delta_{1/2}$	
$\phi(z) = \sqrt{z}/2$	$d\sigma_\phi(\lambda) = \frac{1}{4\sqrt{\pi}}\lambda^{-1/2}\, d\lambda$	for $\lambda > 0$
$\phi(z) = z^\alpha$	$d\sigma_\phi(\lambda) = \frac{\alpha}{\Gamma(1-\alpha)}\lambda^{\alpha}\, d\lambda$	for $\lambda > 0;\ 0 < \alpha < 1$
$\phi(z) = \log(1+z)$	$d\sigma_\phi(\lambda) = \exp(-\lambda)\, d\lambda$	for $\lambda \geq 0$
$\phi(z) = z/(1+z)$	$d\sigma_\phi(\lambda) = \lambda\exp(-\lambda)\, d\lambda$	for $\lambda \geq 0$

Based on the empirical counterparts η_{F_m} and η_{G_n} the test statistic has the representation

$$T^{\phi}_{m,n} = \frac{mn}{m+n}\left[\int |\eta_{F_m}(z) - \eta_{G_n}(z)|^2\, d\tau_\phi(z) + 2\sigma_\phi(\{0\})|\overline{X} - \overline{Y}|^2\right]. \tag{2.4}$$

Examples of functions $\phi : [0,\infty) \to [0,\infty)$ with $\phi(0) = 0$ and completely monotone derivative on $(0,\infty)$ are given in Table 1. The function $\phi(z) = \sqrt{z}/2$ leads to the special Cramér test statistic $T_{m,n}$ suggested by Baringhaus and Franz (2004), Klebanov et al. (2006), Szabo et al. (2002) and Székely (2004). Bahr (1996) proposes the test statistic $[mn/(m+n)]\int|\varphi_{F_m}(z)-\varphi_{G_n}(z)|^2\, dN_d(0, I_d)(z)$, which can be derived from the general class $T^{\phi}_{m,n}$ by choosing $\phi(z) = 1 - \exp(-z/2)$. Szabo et al. (2002) consider the statistics corresponding to $\phi(z) = z^\alpha$. The test statistics obtained by $\phi(z) = \log(1 + z)$ and $\phi(z) = z/(1 + z)$ do not seem to have been studied elsewhere.

3. The Limiting Null Distribution

Let Φ_0 be the familiy of functions $\phi \in \Phi$, the associated measures σ_ϕ of which satisfy $\sigma_\phi(\{0\}) = 0$. Note, that all entries in Table 1 belong to Φ_0. Then we have, compared to (2.4), the somewhat simpler representation

$$T^{\phi}_{m,n} = \frac{mn}{m+n}\int |\eta_{F_m}(z) - \eta_{G_n}(z)|^2\, d\tau_\phi(z).$$

Putting $\operatorname{cosi}(u) := \cos(u) + \sin(u),\ u \in \mathbb{R}$,

$$\mathbb{J}_{\{X_1,\ldots,X_m\};F}(z) := \frac{1}{\sqrt{m}}\sum_{k=1}^{m}\mathbb{X}_k(z),\ \ z \in \mathbb{R}^d$$

with $\mathbb{X}_k(z) := \operatorname{cosi}(X_k'z) - \eta_F(z),\ z \in \mathbb{R}^d$, and $\mathbb{J}_{\{Y_1,\ldots,Y_n\};G}(z) := (1/\sqrt{n})\sum_{k=1}^{n}\mathbb{Y}_k(z)$, $z \in \mathbb{R}^d$ with $\mathbb{Y}_k(z) := \operatorname{cosi}(Y_k'z) - \eta_G(z),\ z \in \mathbb{R}^d$, we can write in the case $F = G$,

$$T^{\phi}_{m,n} = \int\left[\sqrt{\frac{n}{m+n}}\mathbb{J}_{\{X_1,\ldots,X_m\};F}(z) - \sqrt{\frac{m}{m+n}}\mathbb{J}_{\{Y_1,\ldots,Y_n\};G}(z)\right]^2 d\tau_\phi(z). \tag{3.1}$$

The $\mathbb{X}_k$ and $\mathbb{Y}_k$ are centered random elements in the separable Hilbert space $\mathcal{H} = L_2(\mathbb{R}^d, \mathcal{B}^d, \tau_\phi)$. In what follows, only the limiting behavior of the $\mathcal{H}$-valued random elements $\mathbb{J}_{\{X_1,\ldots,X_m\};F}$ as $m \to \infty$ is discussed; the arguments for $\mathbb{J}_{\{Y_1,\ldots,Y_n\};G}$ are the same. Let '$\xrightarrow{\mathcal{D}}$' denote convergence in distribution. Due to

$$\mathrm{E}\|\mathbb{X}_k\|^2 = \int \mathrm{E}\left[1 - \cos\left((X_1 - X_2)'z\right)\right] d\tau_\phi(z) = \mathrm{E}\phi(|X_1 - X_2|^2) < \infty, \quad (3.2)$$

the Central Limit Theorem in Hilbert spaces, cf., Ledoux and Talagrand (1991), Theorem 10.5, applies to get that

$$\mathbb{J}_{\{X_1,\ldots,X_m\};F} \xrightarrow[m\to\infty]{\mathcal{D}} \mathbb{Z},$$

where $\mathbb{Z}$ is a centered Gaussian random element in $\mathcal{H}$ that has the covariance operator

$$C : \begin{array}{l} \mathcal{H} \to \mathcal{H} \quad , \\ e \mapsto \displaystyle\int c(\,\cdot\,, z) e(z)\ d\tau_\phi(z), \end{array}$$

$$c(s,t) = \int \operatorname{cosi}(s'x)\operatorname{cosi}(t'x)\, dF(x) - \eta_F(s)\eta_F(t),\ \ s,t \in \mathbb{R}^d. \quad (3.3)$$

In the case when the null hypothesis is true also $\mathbb{J}_{\{Y_1,\ldots,Y_n\};F} \xrightarrow[n\to\infty]{\mathcal{D}} \mathbb{Z}$. Being a 'root convex combination' of two independent random elements converging in distribution to two independent copies of $\mathbb{Z}$, it follows that

$$\sqrt{\frac{n}{m+n}}\mathbb{J}_{\{X_1,\ldots,X_m\};F} - \sqrt{\frac{m}{m+n}}\mathbb{J}_{\{Y_1,\ldots,Y_n\};G} \xrightarrow[m,n\to\infty]{\mathcal{D}} \mathbb{Z}. \quad (3.4)$$

The Continuous Mapping Theorem yields that $T^\phi_{m,n}$ converges in distribution to the squared $\mathcal{H}$-norm of $\mathbb{Z}$. From the Karhunen-Loève-expansion of $\mathbb{Z}$ we deduce that the squared $\mathcal{H}$-norm of $\mathbb{Z}$ has the same distribution as $\sum_\sigma \lambda_\sigma Z_\sigma^2$, where $(Z_\sigma)_\sigma$ is a sequence of independent unit normal variables and $(\lambda_\sigma)_\sigma$ is the sequence of the positive eigenvalues of the operator C associated with the covariance function (3.3). Theorem 3.1 summarizes the result.

Theorem 3.1. *In the case $F = G \in \mathcal{F}_d(\phi)$ it holds that*

$$T^\phi_{m,n} \xrightarrow[m,n\to\infty]{\mathcal{D}} \sum_\sigma \lambda_\sigma Z_\sigma^2. \quad (3.5)$$

If $F \in \mathcal{F}_d(\phi)$ and $\int \phi^2(|x|^2)\,dF(x) < \infty$ we consider the integral operator

$$H : \begin{array}{ccc} L_2(\mathbb{R}^d, \mathcal{B}^d, F) & \to & L_2(\mathbb{R}^d, \mathcal{B}^d, F), \\ f & \mapsto & \displaystyle\int h(\,\cdot\,, x)\; f(x)\; dF(x), \end{array} \tag{3.6}$$

$$\begin{aligned} h(x_1, x_2) := \iint \Big[& \phi(|x_1 - y_2|^2) + \phi(|x_2 - y_1|^2) \\ & -\phi(|x_1 - x_2|^2) - \phi(|y_1 - y_2|^2)\Big]\, dF(y_1)\; dF(y_2), \end{aligned} \tag{3.7}$$

$x_1, x_2 \in \mathbb{R}^d$. The eigenvalues λ_σ appearing in the limit distribution (3.5) can be found by considering the eigenvalues of the operator H.

Lemma 3.2. *Let $\phi \in \Phi_0$ and $F \in \mathcal{F}_d(\phi)$ with $\int \phi^2(|x|^2)\,dF(x) < \infty$. Then the operator H is positive and of trace class. Additionally, H has the same positive eigenvalues with the same multiplicities as the operator C.*

Proof. Let $f \in L_2(\mathbb{R}^d, \mathcal{B}^d, F)$. From

$$\begin{aligned} &\iint h(x_1, x_2) f(x_1) f(x_2)\, dF(x_1) dF(x_2) \\ &\quad = \int \left[\int [\mathrm{cosi}(x't) - \eta_F(t)] f(x)\, dF(x) \right]^2 d\tau_\phi(t), \end{aligned}$$

we deduce that the operator H is positive. Let $(f_{\tilde\sigma})_{\tilde\sigma}$ be the eigenfunctions to the non-vanishing positive eigenvalues $(\tilde\lambda_{\tilde\sigma})_{\tilde\sigma}$ of H. Then, the $e_{\tilde\sigma}$ with

$$e_{\tilde\sigma}(t) := \frac{1}{\sqrt{\tilde\lambda_{\tilde\sigma}}} \int [\mathrm{cosi}(t'x) - \eta_F(t)] f_{\tilde\sigma}(x)\; dF(x), \;\; t \in \mathbb{R}^d,$$

are orthonormal eigenfunctions of C with the eigenvalues $\tilde\lambda_{\tilde\sigma}$. The orthonormality follows from the identity

$$\int [\mathrm{cosi}(t'x) - \eta_F(t)][\mathrm{cosi}(t'y) - \eta_F(t)]\, d\tau_\phi(t) = h(x, y).$$

The $e_{\tilde\sigma}$ are eigenfunctions of C since

$$\begin{aligned} &\int c(s,t) e_{\tilde\sigma}(t)\; d\tau_\phi(t) \\ &= \frac{1}{\sqrt{\tilde\lambda_{\tilde\sigma}}} \iiint [\mathrm{cosi}(s'x) - \eta_F(s)][\mathrm{cosi}(t'x) - \eta_F(t)] \\ &\qquad\qquad [\mathrm{cosi}(t'y) - \eta_F(t)] f_\sigma(y)\; dF(x)\; dF(y)\; d\tau_\phi(t) \\ &= \frac{1}{\sqrt{\tilde\lambda_{\tilde\sigma}}} \int \left[[\mathrm{cosi}(s'x) - \eta_F(s)] \int h(x,y) f_\sigma(y)\, dF(y) \right] dF(x) \end{aligned}$$

$$
\begin{aligned}
&= \sqrt{\tilde\lambda_{\tilde\sigma}} \int [\mathrm{cosi}(s'x) - \eta_F(s)] f_{\tilde\sigma}(x)\, dF(x) \\
&= \tilde\lambda_{\tilde\sigma} e_{\tilde\sigma}(s).
\end{aligned}
$$

Hence, the positive eigenvalues of H are eigenvalues of C. As a covariance operator, C is of trace class. From

$$
\begin{aligned}
\mathrm{trace}(C) = \sum_\sigma \lambda_\sigma = \int c(t,t)\; d\tau_\phi(t) &= \int \phi(|x-y|^2)\, dF \otimes F(x,y) \\
&= \int h(x,x)\, dF(x) = \mathrm{trace}(H) = \sum_{\tilde\sigma} \tilde\lambda_{\tilde\sigma}
\end{aligned}
$$

see, e.g., Brislawn (1991), we obtain that the operators C and H have the same positive eigenvalues with same multiplicities.

4. Approximation of the Limit Distribution

Due to the fact that the limit distribution is not distribution-free, methods are needed to give an approximation of the critical value. One classical method commonly applied to this kind of problem is the Monte Carlo bootstrap procedure described in Section 1. Using a central limit theorem for triangular schemes in Hilbert spaces, see e.g., Kundu, Majundar, and Mukherjee (2000), one can show that this method works. Henze, Klar, and Meintanis (2003) proceed in this way to prove a permutational limit theorem for a related test statistic. Instead, we suggest using the 'bootstrap in the limit method' described above. To show that this works, we first prove a result on the $\mathcal{H}$-norm convergence of operators associated with covariance functions of the form (3.3), which is interesting in its own right.

Lemma 4.1. *Let $\phi \in \Phi_0$, and let $(F_j)_{j=1}^\infty$ be a sequence of distributions on $\mathcal{B}^d$ converging weakly to some distribution F on $\mathcal{B}^d$. Assume that*

$$
\lim_{j\to\infty} \int \phi(|x-y|^2)\, dF_j \otimes F_j(x,y) = \int \phi(|x-y|^2)\, dF \otimes F(x,y). \tag{4.1}
$$

Let C_j be the operator on $\mathcal{H}$ associated with the covariance function c_j obtained when replacing F by F_j in (3.3). Then $C_j \to C$ in the operator norm on $\mathcal{H}$.

Proof. Let $\eta = \eta_F$ and $\eta_j = \eta_{F_j}$ be the sine-cosine transforms of F and F_j, i.e. $\eta(s) = \int \mathrm{cosi}(s'x)\, dF(x),\ \eta_j(s) = \int \mathrm{cosi}(s'x)\, dF_j(x),\ s \in \mathbb{R}^d$. As $j \to \infty$, for each compact subset M in $\mathbb{R}^d \setminus \{0\}$,

$$
\begin{aligned}
&\lim_{j\to\infty} \int \Big[\int_M |\mathrm{cosi}(s'x) - \eta_j(s)|^2\, d\tau_\phi(s) \Big]\, dF_j(x) \\
&\quad = \int \Big[\int_M |\mathrm{cosi}(s'x) - \eta(s)|^2\, d\tau_\phi(s) \Big]\, dF(x),
\end{aligned} \tag{4.2}
$$

$$\begin{aligned}\lim_{j\to\infty}\int\Big[\int_{M^c}&|\mathrm{cosi}(s'x)-\eta_j(s)|^2\,d\tau_\phi(s)\Big]\,dF_j(x)\\ =\int\Big[\int_{M^c}&|\mathrm{cosi}(s'x)-\eta(s)|^2\,d\tau_\phi(s)\Big]\,dF(x).\end{aligned}\tag{4.3}$$

Due to

$$\begin{aligned}\int\phi(|x-y|^2)\,dF_j\otimes F_j(x,y)&=\int\Big[\int_M|\mathrm{cosi}(s'x)-\eta_j(s)|^2\,d\tau_\phi(s)\Big]\,dF_j(x)\\&\quad+\int\Big[\int_{M^c}|\mathrm{cosi}(s'x)-\eta_j(s)|^2\,d\tau_\phi(s)\Big]\,dF_j(x),\\ \int\phi(|x-y|^2)\,dF\otimes F(x,y)&=\int\Big[\int_M|\mathrm{cosi}(s'x)-\eta(s)|^2\,d\tau_\phi(s)\Big]\,dF(x)\\&\quad+\int\Big[\int_{M^c}|\mathrm{cosi}(s'x)-\eta(s)|^2\,d\tau_\phi(s)\Big]\,dF(x),\end{aligned}$$

and (4.1), (4.2) implies (4.3). Since $h_j(x)=\int_M|\mathrm{cosi}(s'x)-\eta_j(s)|^2\,d\tau_\phi(s),\ x\in\mathbb{R}^d,\ j\in\mathbb{N}$, is a sequence of uniformly bounded continuous functions converging uniformly to $h(x)=\int_M|\mathrm{cosi}(s'x)-\eta(s)|^2\,d\tau_\phi(s),\ x\in\mathbb{R}^d$, as $j\to\infty$, (4.2) follows from the weak convergence of the F_j to F. Let $f\in\mathcal{H}$ with $\|f\|\le 1$. We have

$$\begin{aligned}\|(C_j-C)f\|^2=&\int_M\Big[\int[c_j(s,t)-c(s,t)]f(s)\,d\tau_\phi(s)\Big]^2\,d\tau_\phi(t)\\&+\int_{M^c}\Big[\int[c_j(s,t)-c(s,t)]f(s)\,d\tau_\phi(s)\Big]^2\,d\tau_\phi(t).\end{aligned}\tag{4.4}$$

The second term on the right hand side of (4.4) is bounded from above by

$$2\Big\{\int_{M^c}\Big[\int c_j(s,t)f(s)\,d\tau_\phi(s)\Big]^2\,d\tau_\phi(t)+\int_{M^c}\Big[\int c(s,t)f(s)\,d\tau_\phi(s)\Big]^2\,d\tau_\phi(t)\Big\}.$$

We use Fubini's Theorem, the Cauchy-Schwarz inequality, and $\|f\|\le 1$ to get

$$\begin{aligned}&\int_{M^c}\Big[\int c_j(s,t)f(s)\,d\tau_\phi(s)\Big]^2\,d\tau_\phi(t)\\ &=\int_{M^c}\Big[\int\Big\{\int\big(\mathrm{cosi}(s'x)-\eta_j(s)\big)\big(\mathrm{cosi}(t'x)-\eta_j(t)\big)\,dF_j(x)\Big\}f(s)\,d\tau_\phi(s)\Big]^2\,d\tau_\phi(t)\\ &=\int_{M^c}\Big[\int\Big\{\big(\mathrm{cosi}(t'x)-\eta_j(t)\big)\int\big(\mathrm{cosi}(s'x)-\eta_j(s)\big)f(s)\,d\tau_\phi(s)\Big\}\,dF_j(x)\Big]^2\,d\tau_\phi(t)\\ &\le\int_{M^c}\Big[\int\Big\{|\mathrm{cosi}(t'x)-\eta_j(t)|\Big(\int|\mathrm{cosi}(s'x)-\eta_j(s)|^2\,d\tau_\phi(s)\Big)^{1/2}\Big\}\,dF_j(x)\Big]^2\,d\tau_\phi(t)\end{aligned}$$

$$\leq \int \Big[\int_{M^c} |\mathrm{cosi}(t'x) - \eta_j(t)|^2 \, d\tau_\phi(t)\Big] dF_j(x) \int \Big[\int |\mathrm{cosi}(s'x) - \eta_j(s)|^2 d\tau_\phi(s)\Big] dF_j(x)$$
$$= \int \Big[\int_{M^c} |\mathrm{cosi}(t'x) - \eta_j(t)|^2 \, d\tau_\phi(t)\Big] dF_j(x) \int \phi(|x-y|^2) dF_j \otimes F_j(x,y).$$

In the same way we obtain

$$\int_{M^c} \Big[\int c(s,t) f(s) \, d\tau_\phi(s)\Big]^2 d\tau_\phi(t)$$
$$\leq \int \Big[\int_{M^c} |\mathrm{cosi}(t'x) - \eta(t)|^2 \, d\tau_\phi(t)\Big] \, dF(x) \int \phi(|x-y|^2) \, dF \otimes F(x,y).$$

Given some $\epsilon > 0$, we choose a compact $M \subset \mathbb{R}^d \setminus \{0\}$ such that

$$\int \Big[\int_{M^c} |\mathrm{cosi}(t'x) - \eta(t)|^2 \, d\tau_\phi(t)\Big] \, dF(x) \int \phi(|x-y|^2) \, dF \otimes F(x,y) < \frac{\epsilon}{12}.$$

Then

$$\limsup_{j\to\infty} \sup_{f\in\mathcal{H}, \|f\|\leq 1} \int_{M^c} \Big[\int [c_j(s,t) - c(s,t)] f(s) \, d\tau_\phi(s)\Big]^2 d\tau_\phi(t) \leq \frac{\epsilon}{3}. \tag{4.5}$$

The first term on the right hand side of (4.4) is bounded from above by

$$2\int_M \Big[\int_M [c_j(s,t) - c(s,t)]^2 f(s) \, d\tau_\phi(s)\Big]^2 d\tau_\phi(t)$$
$$+ 2\int_M \Big[\int_{M^c} [c_j(s,t) - c(s,t)]^2 f(s) \, d\tau_\phi(s)\Big]^2 d\tau_\phi(t).$$

Applying the Cauchy-Schwarz inequality, we get

$$\int_M \Big[\int_M [c_j(s,t) - c(s,t)] f(s) \, d\tau_\phi(s)\Big]^2 d\tau_\phi(t)$$
$$\leq \int_M \Big[\int_M [c_j(s,t) - c(s,t)]^2 \, d\tau_\phi(s)\Big] \, d\tau_\phi(t).$$

Since, as $j \to \infty$, the functions c_j converge uniformly on compact subsets of $\mathbb{R}^d \times \mathbb{R}^d$ to c, it follows that

$$\limsup_{j\to\infty} \sup_{f\in\mathcal{H}, \|f\|\leq 1} \int_M \Big[\int_M [c_j(s,t) - c(s,t)] f(s) \, d\tau_\phi(s)\Big]^2 d\tau_\phi(t) = 0.$$

The same arguments applied to achieve (4.5) also yield

$$\limsup_{j\to\infty} \sup_{f\in\mathcal{H}, \|f\|\leq 1} \int_M \Big[\int_{M^c} [c_j(s,t) - c(s,t)] f(s) \, d\tau_\phi(s)\Big]^2 d\tau_\phi(t) \leq \frac{\epsilon}{3}.$$

Summarizing, $\limsup_{j\to\infty}\sup_{f\in\mathcal{H},\|f\|\le 1}\|(C_j-C)f\|^2\le\epsilon$, which proves the desired assertion, since $\epsilon>0$ can be chosen arbitrarily.

Theorem 4.2. *Given the assumptions of Lemma* 4.1*, let $\mathbb{Z},\mathbb{Z}_j, j\in\mathbb{N}$, be centered Gaussian random elements in $\mathcal{H}$ with covariance operators $C, C_j, j\in\mathbb{N}$. Then, as $j\to\infty$, $\mathbb{Z}_j\xrightarrow{\mathcal{D}}\mathbb{Z}$.*

Proof. Let $\langle\cdot,\cdot\rangle$ denote the inner product on $\mathcal{H}$. By Lemma 4.1 the characteristic functionals $\hat\mu_j(f)=\exp\bigl(-\langle C_jf,f\rangle/2\bigr)$, $f\in\mathcal{H}$, of the distributions of the $\mathbb{Z}_j$ converge uniformly on bounded spheres to the characteristic functional $\hat\mu(f)=\exp\bigl(-\langle Cf,f\rangle/2\bigr)$, $f\in\mathcal{H}$. Let $(e_k)_k$ be an orthonormal basis of $\mathcal{H}$. Being covariance operators, the C, C_j are of trace class. Due to

$$\begin{aligned}\operatorname{trace}(C)&=\int c(s,s)\,d\tau_\phi(s)=\int\phi(|x-y|^2)\,dF\otimes F(x,y),\\ \operatorname{trace}(C_j)&=\int c_j(s,s)\,d\tau_\phi(s)=\int\phi(|x-y|^2)\,dF_j\otimes F_j(x,y),\end{aligned}$$

we have that $\lim_{j\to\infty}\operatorname{trace}(C_j)=\operatorname{trace}(C)$. For given $\epsilon>0$ choose some $k_0\in\mathbb{N}$ such that

$$0\le\sum_{k>k_0}\langle Ce_k,e_k\rangle=\operatorname{trace}(C)-\sum_{k\le k_0}\langle Ce_k,e_k\rangle\le\frac{\epsilon}{3}.$$

Choose some $j_0\in\mathbb{N}$ such that

$$\operatorname{trace}(C_j)\le\operatorname{trace}(C)+\frac{\epsilon}{3},\quad\text{and}\quad\sum_{k\le k_0}\langle C_je_k,e_k\rangle\ge\sum_{k\le k_0}\langle Ce_k,e_k\rangle-\frac{\epsilon}{3}$$

for each $j\ge j_0$. Then

$$\begin{aligned}0\le\sum_{k>k_0}\langle C_je_k,e_k\rangle&=\operatorname{trace}(C_j)-\sum_{k\le k_0}\langle C_je_k,e_k\rangle\\ &\le\operatorname{trace}(C)-\sum_{k\le k_0}\langle Ce_k,e_k\rangle+\frac{2\epsilon}{3}\\ &=\sum_{k>k_0}\langle Ce_k,e_k\rangle+\frac{2\epsilon}{3}\le\epsilon.\end{aligned}$$

This gives $\lim_{k\to\infty}\sup_{j\in\mathbb{N}}\sum_{\ell>k}\langle C_je_\ell,e_\ell\rangle=0$. Thus, we have shown that the sequence of operators C_j is compact. The compactness of the sequence (C_j) and the convergence of the characteristic functionals stated above implies the distributional convergence $\mathbb{Z}_j\xrightarrow{\mathcal{D}}\mathbb{Z}$, see, e.g., Parthasarathy (1967).

Corollary 4.3. *Let $(\lambda_{j,\sigma})_\sigma$ and $(\lambda_\sigma)_\sigma$ be enumerations of the positive eigenvalues of C_j and C, and let $(Z_{j,\sigma})_\sigma$ and $(Z_\sigma)_\sigma$ be sequences of independent unit normal* *variables. Then, as $j \to \infty$,*

$$\sum_{\sigma} \lambda_{j,\sigma} Z_{j,\sigma}^2 \xrightarrow{\mathcal{D}} \sum_{\sigma} \lambda_{\sigma} Z_{\sigma}^2.$$

Proof. From Theorem 4.2 and the Continuous Mapping Theorem it follows that $\|\mathbb{Z}_j\|^2 \xrightarrow{\mathcal{D}} \|\mathbb{Z}\|^2$ as $j \to \infty$. The Karhunen-Loéve expansions of $\mathbb{Z}$ and $\mathbb{Z}_j$ implies that $\|\mathbb{Z}\|^2$ and $\|\mathbb{Z}_j\|^2$ have the same distributions as $\sum_\sigma \lambda_\sigma Z_\sigma^2$ and $\sum_\sigma \lambda_{j,\sigma} Z_{j,\sigma}^2$.

To see that the result of the foregoing lemma applies to the approximation of the limit distribution described above, we need to show that the condition (4.1) is fulfilled almost everywhere if $(F_j)_{j=1}^\infty$ is a certain sequence of empirical distributions.

Theorem 4.4. *Let $X_1, X_2, \ldots, Y_1, Y_2, \ldots$ be independent random d-vectors. Let the X_j be identically distributed with distribution $F \in \mathcal{F}_d(\phi)$, and let the Y_k be identically distributed with distribution $G \in \mathcal{F}_d(\phi)$. Let $(m_j)_{j=1}^\infty$ and $(n_j)_{j=1}^\infty$ be sequences of integers tending to infinity in such a way that* $\lim_{j\to\infty}[m_j/(m_j+n_j)] = \rho \in [0,1]$. *Let $\mu_j = [1/(m_j+n_j)]\Big[\sum_{\ell=1}^{m_j} \delta_{X_\ell} + \sum_{\ell=1}^{n_j} \delta_{Y_\ell}\Big]$ be the empirical distribution of the pooled sample $X_1, \ldots, X_{m_j}, Y_1, \ldots, Y_{n_j}$, and $\mu_\rho = \rho F + (1-\rho) G$. Then, as $j \to \infty$,*

$$\int \phi(|x-y|^2)\, d\mu_j \otimes \mu_j(x,y) \to \int \phi(|x-y|^2)\, d\mu_\rho \otimes \mu_\rho(x,y) \quad \textit{with probability } 1.$$

Proof. Put $\tilde{\mu_j} = [m_j/(m_j+n_j)]F + [n_j/(m_j+n_j)]G$. Let $\eta_{F_j}(z) = (1/m_j)\sum_{\ell=1}^{m_j} \operatorname{cosi}(z'X_\ell)$, $z \in \mathbb{R}^d$, and $\eta_{G_j}(z) = (1/n_j)\sum_{\ell=1}^{n_j} \operatorname{cosi}(z'Y_\ell)$, $z \in \mathbb{R}^d$, be the empirical sine-cosine transforms of the random vectors $X_1, \ldots, X_{m_j}$ and the random vectors $Y_1, \ldots, Y_{n_j}$, respectively. By Minkowski's inequality,

$$\begin{aligned}
d_\phi(\tilde{\mu}_j, \mu_j) &= \Big\{ \int \Big| \frac{m_j}{m_j+n_j} \eta_{F_j}(z) + \frac{n_j}{m_j+n_j} \eta_{G_j}(z) \\
&\qquad - \Big(\frac{m_j}{m_j+n_j} \eta_F(z) + \frac{n_j}{m_j+n_j} \eta_G(z) \Big) \Big|^2 d\tau_\phi(z) \Big\}^{1/2} \\
&= \Big\{ \int \Big| \frac{m_j}{m_j+n_j} \big(\eta_{F_j}(z) - \eta_F(z)\big) + \frac{n_j}{m_j+n_j} \big(\eta_{G_j}(z) - \eta_G(z)\big) \Big|^2 d\tau_\phi(z) \Big\}^{1/2} \\
&\le \frac{m_j}{m_j+n_j} \Big\{ \int |\eta_{F_j}(z) - \eta_F(z)|^2 \, d\tau_\phi(z) \Big\}^{1/2} \\
&\qquad + \frac{n_j}{m_j+n_j} \Big\{ \int |\eta_{G_j}(z) - \eta_G(z)|^2 \, d\tau_\phi(z) \Big\}^{1/2}.
\end{aligned}$$

The Law of Large Numbers in the Hilbert space $\mathcal{H} = L_2(\mathbb{R}^d, \mathcal{B}^d, \tau_\phi)$ yields

$$\int |\eta_{F_j}(z) - \eta_F(z)|^2 \, d\tau_\phi(z) \to 0, \quad \text{and} \quad \int |\eta_{G_j}(z) - \eta_G(z)|^2 \, d\tau_\phi(z) \to 0$$

with probability 1 as $j \to \infty$. Thus,

$$d_\phi(\tilde{\mu}_j, \mu_j) \to 1 \;\text{ with probability 1 as } j \to \infty. \tag{4.6}$$

We have

$$\begin{aligned} d_\phi^2(\tilde{\mu}_j, \mu_j) = 2\int \phi(|x-y|)\, d\tilde{\mu}_j \otimes \mu_j(x,y) - \int \phi(|x-y|)\, d\tilde{\mu}_j \otimes \tilde{\mu}_j(x,y) \\ - \int \phi(|x-y|)\, d\mu_j \otimes \mu_j(x,y). \end{aligned} \tag{4.7}$$

Note that

$$\int \phi(|x-y|)\, d\tilde{\mu}_j \otimes \tilde{\mu}_j(x,y) \to \int \phi(|x-y|)\, d\mu_\rho \otimes \mu_\rho(x,y) \;\text{ as } j \to \infty.$$

Additionally, putting

$$\phi_1(x) = \int \phi(|x-y|^2)\, dF(y), \quad \phi_2(x) = \int \phi(|x-y|^2)\, dG(y), \; x \in \mathbb{R}^d,$$

we have

$$\begin{aligned} &2\int \phi(|x-y|)\, d\tilde{\mu}_j \otimes \mu_j(x,y) \\ &= 2\Big\{ \Big(\frac{m_j}{m_j+n_j}\Big)^2 \frac{1}{m_j} \sum_{\ell=1}^{m_j} \phi_1(X_\ell) + \frac{m_j n_j}{(m_j+n_j)^2} \frac{1}{m_j} \sum_{\ell=1}^{m_j} \phi_2(X_\ell) \\ &\quad + \frac{m_j n_j}{(m_j+n_j)^2} \frac{1}{n_j} \sum_{\ell=1}^{n_j} \phi_1(Y_\ell) + \Big(\frac{n_j}{m_j+n_j}\Big)^2 \frac{1}{n_j} \sum_{\ell=1}^{n_j} \phi_2(Y_\ell) \Big\}. \end{aligned}$$

From $\mathrm{E}\phi_1(X_1) = \mathrm{E}\phi(|X_1 - X_2|^2)$, $\mathrm{E}\phi_2(X_1) = \mathrm{E}\phi(|X_1 - Y_1|^2) = \mathrm{E}\phi_1(Y_1)$, and $\mathrm{E}\phi_2(Y_1) = \mathrm{E}\phi(|Y_1 - Y_2|^2)$, we get by using the Law of Large Numbers, that

$$2\int \phi(|x-y|)\, d\tilde{\mu}_j \otimes \mu_j(x,y) \to 2\int \phi(|x-y|^2)\, d\mu_\rho \otimes \mu_\rho(x,y) \tag{4.8}$$

with probability 1 as $j \to \infty$. Combining (4.6), (4.7), and (4.8) yields the assertion.

In what follows we adopt the assumptions of Theorem 4.4. Let C_j be the operator associated with the covariance function c_j obtained when replacing F

by the empirical distribution μ_j of the pooled sample $X_1, \ldots, X_{m_j}, Y_1, \ldots, Y_{n_j}$. Let $\mathcal{L}(T^\phi|F)$ denote the limiting null distribution of the test statistic $T^\phi_{m,n}$ in the case when F is the common distribution of the sample variables X_k and Y_ℓ. Lemma 4.1, Theorem 4.2, and Theorem 4.4 justify approximating $\mathcal{L}(T^\phi|F)$ by $\mathcal{L}(T^\phi|\mu_j)$. For given $\alpha \in (0,1)$ let $c_{j,\alpha}$ be the $(1-\alpha)$-quantile of $\mathcal{L}(T^\phi|\mu_j)$.

Theorem 4.5. *Let the conditions of Theorem* 4.4 *be satisfied. If* $F = G \in \mathcal{F}_d(\phi)$, *and* F *non-degenerate, then* $\lim_{j\to\infty} P\big(T^\phi_{m_j,n_j} > c_{j,\alpha}\big) = \alpha$. *For all distributions* $F, G \in \mathcal{F}_d(\phi), F \neq G$, *it holds that* $\lim_{j\to\infty} P\big(T^\phi_{m_j,n_j} > c_{j,\alpha}\big) = 1$.

Proof. For each $F, G \in \mathcal{F}_d(\phi)$, Lemma 4.1, Theorem 4.2, and Theorem 4.4 yield, as $j \to \infty$, the weak convergence of $\mathcal{L}(T^\phi|\mu_j)$ to $\mathcal{L}(T^\phi|\mu_\rho)$ with probability 1. This proves the first assertion of the theorem. The Law of Large Numbers for independent and identically distributed random elements in the Hilbert space $\mathcal{H}$ gives that $[(m_j+n_j)/m_jn_j]T^\phi_{m_j,n_j} \to d^2_\phi(F,G)$ with probability 1. Since $d^2_\phi(F,G) > 0$ for $F, G \in \mathcal{F}_d(\phi)$, $F \neq G$, the second assertion of the theorem follows.

To get the quantiles of $\mathcal{L}(T^\phi|\mu_j)$ we need to calculate the positive eigenvalues of the operator C_j. For this purpose we consider the operator H_j defined as in (3.6) with F replaced by μ_j. Due to Lemma 3.2, H_j and C_j have the same positive eigenvalues. Since μ_j has finite support the integral equation

$$\int h_j(x,y) f(y)\, d\mu_j(y) = \lambda f(x), \;\; f \in L_2(\mathbb{R}^d, \mathcal{B}^d, \mu_j),$$

reduces to a matrix eigenvalue problem that can be solved easily. Here, h_j denotes the kernel function (3.7) with F replaced by μ_j. The Fourier transform of $\mathcal{L}(T^\phi|\mu_j)$ is given by

$$\varphi_{\mathcal{L}(T^\phi|\mu_j)}(t) = \prod_\sigma \big(1 - 2\mathrm{i}t\lambda_{j,\sigma}\big)^{-1/2}, \qquad t \in \mathbb{R}.$$

The inverse fast Fourier transform and the inversion formula of Gurland (1948) can be used to calculate the $(1-\alpha)$-quantile of $\mathcal{L}(T^\phi|\mu_j)$. In fact, the distribution of $\mathcal{L}(T^\phi|\mu_j)$ is the distribution of a special quadratic form in normal variables. The problem of computing the distribution of quadratic forms in normal variables has been studied by various authors. The interested reader is referred to the papers of Imhof (1961) and Martynov (1975), and to the references therein.

5. Efficiency

A theoretical comparison of the new two-sample tests for different ϕ is easily done by using the concept of approximate slopes proposed by Bahadur (1960).

The ratio of approximate slopes of two sequences of test statistics is called their approximate Bahadur efficiency. These are of limited use as means of comparing tests; for example, monotone transformations of the test statistics can lead to different approximate slopes. Nevertheless, in typical cases approximate Bahadur efficiencies coincide locally with Pitman efficiencies, see Wieand (1976) and Kallenberg and Koning (1995). In what follows, we confine ourselves to an informal approach giving a first rough impression of the performance of the new tests for normal location alternatives, normal scale alternatives, and Lehmann's contaminated alternative.

For simplicity, we deal solely with the case $d = 1$. As in Theorem 4.4 we start with sample sizes $m = m_j$ and $n = n_j$, where $(m_j)_{j=1}^{\infty}$ and $(n_j)_{j=1}^{\infty}$ are sequences of integers tending to infinity in such a way that $\lim_{j\to\infty}[m_j/(m_j + n_j)] = \rho \in (0, 1)$. To define approximate slopes of the test statistics $T_{m,n}^{\phi}$ we modify the definition given by Wieand (1974) for two-sample problems: we replace his 'norming' factor $\sqrt{m+n}$ by the factor $\sqrt{mn/(m+n)}$. Then the factor $\rho(1-\rho)$ does not appear in the expressions for the approximate slopes.

Consider the simple testing problem of the hypothesis of a common fixed distribution F, and the alternatives of distributions F and $G \neq F$. It is assumed that F and G belong to $\mathcal{F}_d(\phi)$ for all ϕ under consideration. Regard $\big(\big[T_{m,n}^{\phi}\big]^{1/2}\big)_{m,n}$ as a sequence of test statistics. Denote by $\lambda(\phi, F)$ the largest eigenvalue of the covariance operator C associated with F and ϕ by (3.3). Adjusting Bahadur's definition to the two-sample case, it is easily seen, see e.g., Koziol (1986), that $\big(\big[T_{m,n}^{\phi}\big]^{1/2}\big)_{m,n}$ is a standard sequence with the approximate slope

$$s(\phi, F, G) = \frac{d_{\phi}^2(F, G)}{\lambda(\phi, F)}.$$

We are interested in power performance as G tends to F. To this end, we assume that G belongs to some parametric family $\{G_\theta; \theta \in \Theta\}$ of distributions, where Θ is an open subset of $\mathbb{R} \setminus \{0\}$ and $G_\theta \to F$ weakly as $\theta \to 0$. In what follows the parametric families considered have the property that there is some real positive $a(\phi, F)$ such that

$$s(\phi, F) = \lim_{\theta\to 0} \frac{s(\phi, F, G_\theta)}{\theta^2} = \frac{a(\phi, F)}{\lambda(\phi, F)},$$

the limiting approximate Bahadur slope. The ratios $s(\phi_1, F)/s(\phi_2, F)$ for different ϕ_1, ϕ_2 are the limiting approximate Bahadur efficiencies. We give limiting approximate Bahadur slopes for the function $\phi(z) = 1 - \exp(-z/2)$ leading to Bahr's test, the function $\phi(z) = \sqrt{z}/2$ leading to the Cramér test, and the functions $\phi(z) = \log(1+z)$ and $\phi(z) = z/(1+z)$. At first we consider normal location

Table 2. Limiting approximate slopes for normal location alternatives.

$\phi(z)$	$\sqrt{z}/2$	$1-\exp(-z/2)$	$\log(1+z)$	$z/(1+z)$
$\lambda\big(\phi, N(0,1)\big)$	0.29727	0.23607	0.49493	0.20361
$s\big(\phi, N(0,1)\big)$	0.949	0.815	0.918	0.782

test	t	WMW	CvM	KS	NN
slope	1	0.955	0.907	0.637	0

Table 3. Limiting approximate slopes for normal scale alternatives.

$\phi(z)$	$\sqrt{z}/2$	$1-\exp(-z/2)$	$\log(1+z)$	$z/(1+z)$
$\lambda\big(\phi, N(0,1)\big)$	0.29727	0.23607	0.49493	0.20361
$s\big(\phi, N(0,1)\big)$	0.474	0.815	0.596	1.234

test	F	Mood	CvM	KS	NN
slope	2	1.520	0.302	0.234	0

alternatives and normal scale alternatives. Here, F is $\mathrm{N}(0,1)$, the location alternatives are $G_\theta = \mathrm{N}(\theta, 1)$, the scale alternatives are $G_\theta = \mathrm{N}\big(0, (1+\theta)^2\big)$ with $\theta \in (-1, \infty), \theta \neq 0$. Table 2 and Table 3 show the limiting approximate Bahadur slopes and the largest eigenvalues $\lambda\big(\phi, \mathrm{N}(0,1)\big)$. Except for $\phi(z) = 1-\exp(-z/2)$ where exact results are available, see Baringhaus (1996), the eigenvalues are computed by numerical methods. The last rows of Table 3 and Table 4 give the limiting approximate slopes of some competing tests.

For normal location alternatives the competitors considered are the two-sided t-test, the two-sided Wilcoxon-Mann-Whitney test (WMW), the Cramér-von Mises test (CvM), the Kolmogorov-Smirnov test (KS), and the nearest neighbor test (NN) of Henze (1984). For normal scale alternatives, instead of the t-test and the Wilcoxon-Mann-Whitney test, the two-sided F-test and the Mood rank test, see Mood (1954), are chosen.

The limiting approximate slopes for the t-test, F-test, Cramér-von Mises test, and the Kolmogorov-Smirnov test are obtained from Wieand (1976), that of the Wilcoxon-Mann-Whitney test from Hollander (1967). The limiting approximate slope of the Mood test is easily derived from the work of Mood (1954). Using the results given by Henze (1984), it is easily verified that the limiting approximate Bahadur slope of the nearest neighbor test is 0.

Lehmann's contaminated alternative considered here is $G_\theta = (1-\theta)U + \theta U^2,\ \theta \in (0,1)$, where $U = \mathrm{U}[0,1]$ denotes the uniform distribution on the unit interval $[0,1]$, and U^2 is the distribution of the maximum of two independent random variables uniformly distributed on the unit interval. The limiting approximate Bahadur slopes are shown in Table 4 where the last row gives the limiting approximate slopes of some competing tests: the Wilcoxon-Mann-Whitney test,

Table 4. Limiting approximate slopes for Lehmann's contaminated alternative.

$\phi(z)$	$\sqrt{z}/2$	$1-\exp(-z/2)$	$\log(1+z)$	$z/(1+z)$
$\lambda\big(\phi, \mathrm{U}(0,1)\big)$	0.1013	0.0721	0.1313	0.1058
$s\big(\phi, \mathrm{U}(0,1)\big)$	0.329	0.333	0.333	0.331

test	WMW	CvM	KS	NN
slope	0.333	0.329	0.250	0

the Cramér-von Mises test, the Kolmogorov-Smirnov test, and the nearest neighbor test of Henze. The limiting approximate slope for the Cramér-von Mises test is clearly the same as that of the Cramér test, that of the Kolmogorov-Smirnov test is obtained from Bahadur (1960), and that of the Wilcoxon-Mann-Whitney test from Hodges and Lehmann (1956). The limiting approximate slope of the Mood test is easily derived from the work of Mood (1954).

To prove the equivalence of limiting approximate Bahadur efficiency and limiting Pitman efficiency, one needs to verify a version of Wieand's condition III*, see Wieand (1974, 1976), adjusted to the two-sample case. For the new tests the condition is as follows.

Wieand's condition III*: There exists some $\theta^* > 0$, such that for each $\varepsilon > 0$ and each $\delta \in (0,1)$, there is a real positive constant C such that for all $\theta \in (-\theta^*, +\theta^*) \setminus \{0\}$ and all $j \in \mathbb{N}$ with $m_j, n_j > C/b^2(\theta)$,

$$P_\theta\left(\left|d_\phi(F_{m_j}, G_{n_j}) - d_\phi(F, G_\theta)\right| < \varepsilon d_\phi(F, G_\theta)\right) > 1 - \delta. \tag{5.1}$$

For the normal location alternatives, normal scale alternatives, and Lehmann's contaminated alternative, it can be shown that

$$\int \phi(|x-y|^2)\, dG_\theta \otimes G_\theta(x,y) \to \int \phi(|x-y|^2)\, dF \otimes F(x,y) \quad \text{as } \theta \to 0.$$

Having such a result Wieand's condition III* is easily verified by applying Markov's inequality and using the fact that d_ϕ is a metric on $\mathcal{F}_d(\phi)$. For details we refer to Franz (2004).

6. Simulation Studies

We have studied the new bootstrap in the limit method. For various univariate and multivariate distributions and for different levels chosen, almost the same behavior is observed. The deviations of the estimated error probabilities from the given levels are satisfactorily small, except for the case of higher dimensions ($d \geq 4$) and small sample sizes ($m, n \leq 20$), see Franz (2004). In the latter cases it is recommended to use the conventional Monte Carlo bootstrap procedure. As

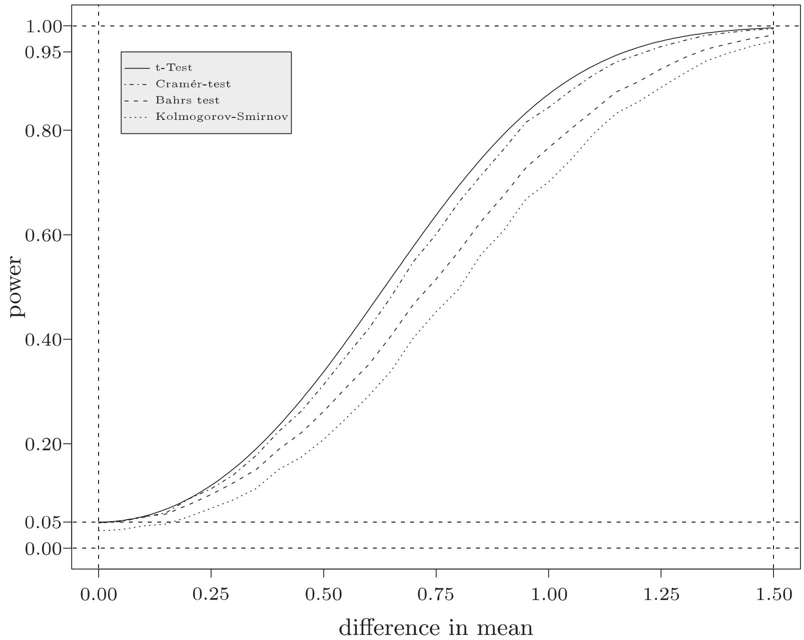

Figure 1. Power values for normal location alternatives, Study (a).

we are interested in the power of the new tests with the critical values obtained by the bootstrap in the limit method, sample sizes $m = n = 20$ were chosen only for some univariate cases, whereas for multivariate problems $m = n = 50$. The significance level was always $\alpha = 0.05$. The first power values shown here are for the three types of univariate alternatives described in Section 5:

(a) normal location alternatives with $F = \mathrm{N}(0,1),\ G = \mathrm{N}(\theta,1)$;

(b) normal scale alternatives with $F = \mathrm{N}(0,1),\ G = \mathrm{N}(0,\sigma^2)$;

(c) Lehmann's contaminated alternative with $F = U = \mathrm{U}[0,1],\ G = (1-\theta)U + \theta U^2$.

In the 3-dimensional case, powers are given

(d) for Lehmann's contaminated alternative $F = U = \mathrm{U}[0,1]^3,\ G = (1-\theta)U + \theta U^2$ where $U = \mathrm{U}[0,1]^3$ denotes the uniform distribution on the 3-dimensional unit cube $[0,1]^3$, and U^2 is the distribution of the random vector the components of which are the maxima of the two corresponding components of two independent copies with distribution U.

We used 10,000 replications. For normal location alternatives, Figure 1 shows the exact power values of the t-test and the empirical power values for the Cramér

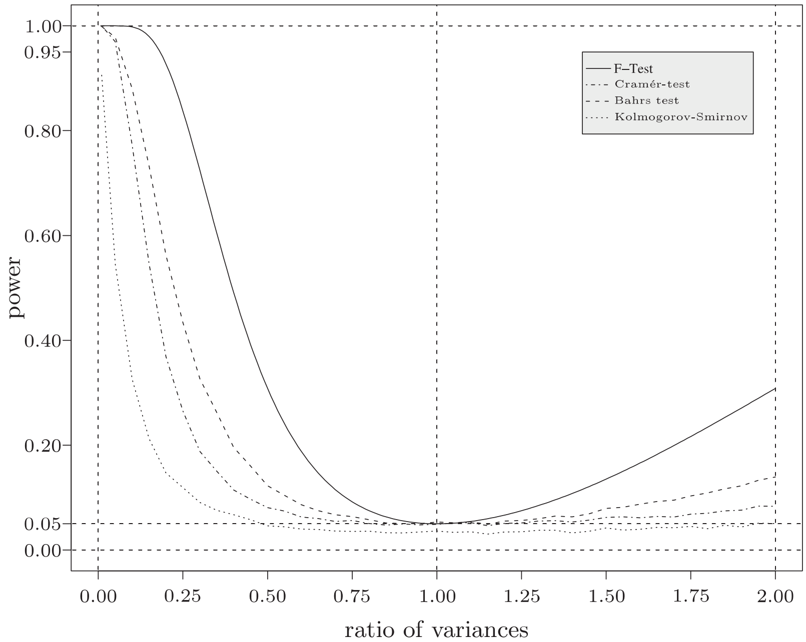

Figure 2. Power values for normal scale alternatives, Study (b).

test, Bahr's test and the Kolmogorov-Smirnov test as functions of the difference in mean θ.

The t-test outperforms its competitors. It is followed by the Cramér test of the new family. The Mann-Whitney test, the Cramér-von Mises test, and the new tests with $\phi(z) = \log(1+z)$ and $\phi(z) = z/(1+z)$ are not shown. The behavior of each of the first three of these tests is similar to that of the Cramér test; the power of the tests with $\phi(z) = z/(1+z)$ lies between Bahr's test and the Kolmogorov-Smirnov test.

For normal scale alternatives, Figure 2 shows the power values as functions of the ratio of variances σ^2 of the underlying distributions F and G. The sample sizes were $m = n = 20$. The F-test clearly outperforms its competitors. It is followed by Bahr's test, which behaves similar to the test with $\phi(z) = z/(1+z)$, not shown. The Cramér test, which performs almost like the test with $\phi(z) = \log(1+z)$, not shown, is still noticeable better than the Kolmogorov-Smirnov test. The latter is slightly worse than the Cramér-von Mises test and Henze's nearest-neighbor test, also not shown.

In the multivariate normal setting of the location or dispersion problem generally the same observations can be made. The Cramér-test and the test with $\phi(z) = \log(1+z)$ do very well for the location problem, whereas Bahr's test and the test with $\phi(z) = z/(1+z)$ are suited for dispersion alternatives. For a detailed description of the results in the multivariate case see Franz (2004).

Table 5. Power values for Lehmann's contaminated alternative, d=1, Study (c).

θ	Cramér	log(1+z)	Bahr	$z/(1+z)$	WMW	KS	CvM	NN
0.0	0.053	0.054	0.054	0.053	0.051	0.038	0.053	0.045
0.2	0.089	0.093	0.094	0.092	0.090	0.066	0.089	0.047
0.4	0.207	0.212	0.213	0.211	0.210	0.153	0.204	0.055
0.6	0.420	0.433	0.433	0.430	0.416	0.326	0.408	0.080
0.8	0.650	0.666	0.667	0.661	0.635	0.521	0.629	0.112
1.0	0.859	0.873	0.875	0.869	0.833	0.738	0.829	0.177

Table 6. Power values for Lehmann's contaminated alternative, d=3, Study (d).

θ	Cramér	log(1+z)	Bahr	$z/(1+z)$	NN
0.0	0.050	0.051	0.050	0.051	0.054
0.2	0.117	0.121	0.121	0.120	0.070
0.4	0.358	0.362	0.361	0.364	0.108
0.6	0.715	0.719	0.718	0.718	0.192
0.8	0.944	0.948	0.948	0.945	0.355
1.0	0.998	0.999	0.998	0.998	0.625

For Lehmann's contaminated alternative the cases $d = 1$ and $d = 3$ were considered and simulation results are shown for $\theta = 0, 0.2, \ldots, 1.0$, and sample sizes $m = n = 50$. Table 5 gives the empirical power values of the competing tests (Cramér, Bahr, new class with $\phi(z) = \log(1 + z)$, $\phi(z) = z/(1 + z)$, WMW, KS, CvM, and NN) in the univariate case. As for the normal location and normal scale alternatives, the new family of tests exhibits a good power performance, comparing favorably with the two-sided Wilcoxon-Mann-Whitney test.

Table 6 shows the simulation results for the multivariate case $d = 3$. The tests of the new class (Cramér, Bahr, the tests with $\phi(z) = \log(1+z)$, $\phi(z) = z/(1+z)$) have comparable power and clearly outperform the test of Henze.

The last simulations presented in the paper involve the multivariate t-distribution $\mathbf{t}_d(\nu)$ with parameter (degrees of freedom) $\nu > 0$, the multivariate logistic distribution $\mathbf{L}_d(\kappa)$ with parameter $\kappa > 0$, the uniform distribution $\mathbf{U}_d(\mu)$ on the unit ball centered at $\mu \in \mathbb{R}^d$, and the multivariate Weibull distribution $\mathbf{W}_d(\beta)$ with independent univariate Weibull components and parameter $\beta > 0$. Only the case $d = 2$ was considered. The pairs of distributions chosen were

(e) $F = N_2(0, I_2)$ and $G = \mathbf{t}_2(\nu)$ with $\nu \in \{2, 5, 10, 20, 50\}$,

(f) $F = \mathbf{L}_2(1)$ and $G = \mathbf{L}_2(\kappa)$ with $\kappa \in \{1/2, 3/4, 1, 3/2, 2\}$,

(g) $F = \mathbf{U}_2(0)$ and $G = \mathbf{U}_2\big((\theta, \theta)'\big)$ (location alternatives) with $\theta = 0.00$, $0.05, \ldots, 0.45$, and

(h) $F = \mathbf{W}_2(1)$ and $G = \mathbf{W}_2(\beta)$ with $\beta = 1.0, 1.2, \ldots, 1.8$.

Table 7. Power values for $\mathbf{t}_2(\nu)$-distributions, Study (e).

ν	Cramér	log(1+z)	Bahr	$z/(1+z)$	NN
2	0.685 (0.672)	0.672 (0.670)	0.384 (0.390)	0.404 (0.409)	0.197
5	0.134 (0.136)	0.136 (0.138)	0.107 (0.109)	0.101 (0.103)	0.072
10	0.067 (0.067)	0.066 (0.067)	0.062 (0.063)	0.061 (0.061)	0.055
20	0.047 (0.049)	0.047 (0.047)	0.049 (0.052)	0.047 (0.048)	0.045
50	0.049 (0.051)	0.050 (0.051)	0.050 (0.052)	0.047 (0.049)	0.044

Table 8. Power values for $\mathbf{L}_2(\kappa)$-distributions, Study (f).

κ	Cramér	log(1+z)	Bahr	$z/(1+z)$	NN
1/2	0.927 (0.926)	0.870 (0.873)	0.498 (0.503)	0.591 (0.593)	0.206
3/4	0.290 (0.291)	0.239 (0.242)	0.115 (0.117)	0.127 (0.130)	0.059
1	0.053 (0.053)	0.049 (0.051)	0.043 (0.045)	0.041 (0.042)	0.040
3/2	0.507 (0.506)	0.442 (0.445)	0.225 (0.227)	0.253 (0.257)	0.089
2	0.940 (0.939)	0.908 (0.908)	0.654 (0.657)	0.715 (0.718)	0.240

Table 9. Power values for uniform $\mathbf{U}_2\big((\theta,\theta)'\big)$ location alternatives, Study (g).

θ	Cramér	log(1+z)	Bahr	$z/(1+z)$	NN
0.00	0.051 (0.052)	0.052 (0.053)	0.052 (0.053)	0.050 (0.051)	0.044
0.05	0.080 (0.080)	0.081 (0.083)	0.080 (0.081)	0.074 (0.075)	0.054
0.10	0.179 (0.183)	0.180 (0.185)	0.173 (0.178)	0.150 (0.152)	0.078
0.15	0.365 (0.370)	0.364 (0.366)	0.347 (0.348)	0.298 (0.302)	0.131
0.20	0.614 (0.618)	0.611 (0.612)	0.582 (0.584)	0.515 (0.518)	0.223
0.25	0.826 (0.829)	0.818 (0.818)	0.792 (0.793)	0.737 (0.739)	0.347
0.30	0.944 (0.945)	0.938 (0.937)	0.920 (0.922)	0.891 (0.893)	0.499
0.35	0.989 (0.989)	0.986 (0.986)	0.979 (0.979)	0.967 (0.968)	0.651
0.40	0.999 (0.999)	0.999 (0.999)	0.997 (0.997)	0.995 (0.995)	0.776
0.45	1.000 (1.000)	1.000 (1.000)	1.000 (1.000)	1.000 (0.999)	0.875

Table 10. Power values for Weibull $\mathbf{W}_2(\beta)$ alternatives, Study (h).

β	Cramér	log(1+z)	Bahr	$z/(1+z)$	NN
1.0	0.046 (0.047)	0.048 (0.048)	0.049 (0.052)	0.046 (0.048)	0.044
1.2	0.079 (0.081)	0.090 (0.091)	0.102 (0.104)	0.101 (0.102)	0.061
1.4	0.189 (0.192)	0.239 (0.243)	0.318 (0.321)	0.313 (0.316)	0.125
1.6	0.421 (0.423)	0.543 (0.543)	0.672 (0.674)	0.667 (0.670)	0.241
1.8	0.712 (0.706)	0.814 (0.814)	0.902 (0.902)	0.903 (0.904)	0.407

The empirical power values of the competing tests (Cramér, Bahr, new class with $\phi(z) = \log(1+z)$, $\phi(z) = z/(1+z)$, and Henze's nearest neighbor test NN) are shown in Tables 8−10. For comparison, the power values of the new class tests, with critical values obtained by the traditional bootstrap Monte Carlo procedure, are shown in parentheses.

7. Recommendations

The package `cramer` for the free statistical software environment **R** available under the address `http://cran.r-project.org` offers a simple way to apply the new tests to given data. p-values or critical values can be obtained either with the bootstrap in the limit or with the traditional bootstrap Monte Carlo procedure. Our experiments with this software show that, on the basis of the default values for numerical accuracy and the number of Monte Carlo samples, one will come to rejection/acceptation of the hypothesis for given data slightly faster with the bootstrap in the limit method as long as $m, n \leq 300$. Unless $d \geq 4$ and $m, n \leq 20$ the deviations of the estimated error probabilities from the given level are satisfactorily small for this method. Propagating the traditional method for the exceptional cases mentioned, we strongly advocate use of the bootstrap in the limit method in all the other cases.

Compared to Baringhaus and Franz (2004) dealing with the Cramér test for testing $H : F = G$ against the general alternative $H : F \neq G$, we are able by choosing a suitable kernel ϕ to gain more flexibility with respect to the power of the test against specific alternatives. We suggest the use of the Cramér test or the test with $\phi(z) = \log(1 + z)$ for location alternatives. Overall, the test with $\phi(z) = \log(1 + z)$ seems to be a good choice.

8. Extension

Consider the c-sample problem, $c \geq 2$. Let $X_{i1}, \ldots, X_{in_i}, i = 1, \ldots, c$, be c independent random samples of sizes $n_1, \ldots, n_c$, where the random variables X_{ik} of the ith sample are independent and identically distributed with the distribution $F_i \in \mathcal{F}_d(\phi)$. Let F_{in_i} be the empirical distribution of the ith sample, put $n = n_1 + \cdots + n_c$, and $\overline{F} = (1/n) \sum_{i=1}^{c} n_i F_{in_i}$. $\overline{F}$ is the empirical distribution of the pooled sample $X_{1n_1}, \ldots, X_{cn_c}$. For treating the testing problem

$$H : F_1 = \cdots = F_c, \; K : F_i \neq F_j \text{ for some } i \neq j,$$

we suggest the test statistic

$$\begin{aligned} T^{\phi}_{n_1,\ldots,n_c} &= \sum_{i=1}^{c} n_i d^2_{\phi}(F_{in_i}, \overline{F}) \\ &= \sum_{i=1}^{c} n_i \int |\varphi_{F_{in_i}}(z) - \varphi_{\overline{F}}(z)|^2 \, d\tau_{\phi}(z) \\ &= \sum_{i=1}^{c} n_i \int |\eta_{F_{in_i}}(z) - \eta_{\overline{F}}(z)|^2 \, d\tau_{\phi}(z). \end{aligned}$$

Denoting by $\mathcal{L}(T^\phi|F)^{\star c-1}$ the $(c-1)$-fold convolution of $\mathcal{L}(T^\phi|F)$, it can be shown that if the hypothesis is true with F the common distribution, $T^\phi_{n_1,\ldots,n_c}$ tends in distribution to $\mathcal{L}(T^\phi|F)^{\star c-1}$ as $\min(n_1,\ldots,n_c) \to \infty$. The approach of Section 4 is adjusted easily to see that $\mathcal{L}(T^\phi|\overline{F})^{\star c-1}$ provides a suitable approximation to $\mathcal{L}(T^\phi|F)^{\star c-1}$. Since the distribution $\mathcal{L}(T^\phi|F)^{\star c-1}$ is of the same type as $\mathcal{L}(T^\phi|F)$, i.e. the weighted sum of squares of independent unit normal variables, the methods for calculating the critical value also carry over. The test rejecting the hypothesis if $T^\phi_{n_1,\ldots,n_c}$ exceeds the critical value is again seen to be consistent, and to be asymptotically of given level α.

References

Bahadur, R. R. (1960). Stochastic comparison of tests. *Ann. Math. Statist.* **31**, 276-295.

Bahr, R. (1996). Ein neuer Test für das mehrdimensionale Zwei-Stichproben-Problem bei allgemeiner Alternative. Doctoral thesis, Universität Hannover.

Baringhaus, L. (1996). Fibonacci numbers, Lucas numbers and integrals of certain Gaussian processes. *Proc. Amer. Math. Soc.* **124**, 3875-3884.

Baringhaus, L. and Franz, C. (2004). On a new multivariate two-sample test. *J. Multivariate Anal.* **88**, 190-206.

Berg, C., Christensen, J. P. R. and Ressel, P. (1984). *Harmonic Analysis on Semigroups.* Springer, New York.

Bickel, P. (1969). A distribution free version of the Smirnov two-sample test in the multivariate case. *Ann. Math. Statist.* **40**, 1-23.

Brislawn, C. (1991). Traceable integral kernels on countably generated measure spaces. *Pacific J. Math.* **150**, 229-240.

Franz, C. (2004). Asymptotische Eigenschaften einer neuen Klasse von nichtparametrischen Zwei-Stichproben-Tests. Doctoral thesis, Universität Hannover.

Guo, K., Hu, S. and Sun, X. (1993). Conditionally positive definite functions and Laplace-Stieltjes integrals. *J. Approx. Theory* **74**, 249-265.

Gurland, J. (1948). Inversion formulae for the distribution of ratios. *Ann. Math. Statist.* **19**, 228-237.

Henze, N. (1984). Über die Anzahl von Zufallspunkten mit typ-gleichem nächsten Nachbarn und einen multivariaten Zwei-Stichproben-Test. *Metrika* **31**, 259-273.

Henze, N. (1988). A multivariate two-sample test based on the number of nearest neighbor type coincidences. *Ann. Statist.* **16**, 772-783.

Henze, N., Klar, B. and Meintanis, S. G. (2003). Invariant tests for symmetry about an unspecified point based on the empirical characteristic function *J. Multivariate Anal.* **87**, 275-297.

Hettmannsperger, T. and McKean, J. (1998). *Robust Nonparametric Statistical Methods.* Wiley, New York.

Hodges, J. L. and Lehmann, E. L. (1956). The efficiency of some nonparametric competitors of the t-test. *Ann. Math. Statist.* **27**, 324-335.

Hollander, M. (1967). Asymptotic efficiency of two nonparametric competitors of Wilcoxon's two sample test. *J. Amer. Statist. Assoc.* **62**, 939-949.

Imhof, J. P. (1961). Computing the distribution of quadratic forms in normal variables. *Biometrika* **48**, 419-426.

Kallenberg, W. C. M. and Koning, A. J. (1995). On Wieand's theorem. *Statist. Probab. Lett.* **25**, 484-493.

Klebanov, L. B. (2005). $\mathfrak{N}$-*Distances and their Applications,* Charles University, Karolinum Press, Prague.

Klebanov, L., Gordon, A., Xiao, Y., Land, H. and Yakovlev, A. (2006). A permutation test motivated by microarray data analysis. *Comput. Statist. Data Anal.* **50**, 3619-3628.

Koziol, J. A. (1986). Relative efficiencies of goodness of fit procedures for assessing univariate normality. *Ann. Inst. Statist. Math.* **38**, 121-132.

Kundu, S., Majundar, S. and Mukherjee, K. (2000). Central limit theorems revisited. *Statist. Probab. Lett.* **47**, 265-275.

Ledoux, M. and Talagrand, M. (1991). *Probability in Banach spaces.* Springer, Berlin.

Martynov, G. V. (1975). Computation of distribution functions of quadratic forms of normally distributed random variables. *Theor. Probab. Appl.* **20**, 782-793.

Mattner, L. (1990). Behandlung einiger Extremalprobleme für Wahrscheinlichkeitsverteilungen. Doctoral thesis, Universität Hannover.

Mattner, L. (1997). Strict definiteness of integrals via complete monotonicity of derivatives. *Trans. Amer. Math. Soc.* **349**, 3321-3342.

Mood, A. (1954). On the asymptotic efficiency of certain nonparametric two-sample tests. *Ann. Math. Statist.* **25**, 514-522.

Parthasarathy, K. R. (1967). *Probability Measures on Metric Spaces.* Academic Press, New York.

Præstgaard, J. (1995). Permutation and bootstrap Kolmogorov-Smirnov tests for the equality of two distributions. *Scand. J. Statist.* **22**, 305-322.

Puri, M. and Sen, P. (1971). *Nonparametric Methods in Multivariate Analysis.* Wiley, New York.

Schilling, M. (1986). Multivariate two-sample tests based on nearest neighbors. *J. Amer. Math. Assoc.* **81**, 799-806.

Schoenberg, I. J. (1938). Metric spaces and completely monotone functions. *Ann. of Math.* **39**, 811-841.

Szabo, A., Boucher, K., Carroll, W., Klebanov, L., Tsodikov, A. and Yakovlev, A. (2002). Variable selection and pattern recognition with gene expression data generated by the microarray technology. *Mathematical Biosciences* **176**, 71-98.

Szabo, A., Boucher, K., Jones, D., Klebanov, L., Tsodikov, A. and Yakovlev, A. (2003). Multivariate exploratory tools for microarray data analysis. *Biostatistics* **4**, 555-567.

Székely, G. (2004). $\mathcal{E}$-statistics: the energy of statistical samples. Preprint.

Wieand, H. S. (1974). A condition under which the Pitman and Bahadur approaches to efficiency coincide. Ph.D. thesis, University of Maryland.

Wieand, H. S. (1976). A condition under which the Pitman and Bahadur approaches to efficiency coincide. *Ann. Statist.* **4**, 1003-1011.

Zinger, A. A., Klebanov, L. B. and Kakosyan, A. V. (1989). Characterization of distributions by mean values of statistics in connection with some probability metrics. In *Stability Problems for Stochastic Models*, 47-55 (in Russian). VNIISI, Moscow. English Translation in: *J. Math. Sci.* **59**, 914-920.

Zuo, Y. and He, X. (2006). On the limiting distributions of multivariate depth-based rank sum statistics and related tests. *Ann. Statist.* **34**, 2879-2896.

Institut für Mathematische Stochastik, Leibniz Universität Hannover, Postfach 60 09, D-30060 Hannover, Germany.

E-mail: lbaring@stochastik.uni-hannover.de

Institut für Mathematische Stochastik, Leibniz Universität Hannover, Postfach 60 09, D-30060 Hannover, Germany.

E-mail: carsten.franz@gmail.com

(Received October 2008; accepted May 2009)