DOI 10.1007/s10958-024-07498-y
Journal of Mathematical Sciences, Vol. 286, No. 2, December, 2024

HÖFFDING'S KERNELS AND PERIODIC COVARIANCE REPRESENTATIONS

Sergey G. Bobkov*
University of Minnesota
127 Vincent Hall, Minneapolis, MN 55455, USA
bobkov@math.umn.edu

Devraj Duggal
University of Minnesota
127 Vincent Hall, Minneapolis, MN 55455, USA
dugga079@umn.edu

UDC 511.3

We start with a brief survey on the Höffding kernels, its properties, related spectral decompositions, and discuss marginal distributions of Höffding measures. In the second part of this note, one dimensional covariance representations are considered over compactly supported probability distributions in the class of periodic smooth functions. Höffding's kernels are used in the construction of mixing measures whose marginals are multiples of given probability distributions, leading to optimal kernels in periodic covariance representations. Bibliography: 16 *titles.*

1 Generalized Höffding Formula

iven two random variables X and Y, the generalized Höffding's covariance formula indicates that for all "regular" functions u and v on the real line

$$\mathrm{cov}\,(u(X), v(Y)) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} u'(x)v'(y)\,H(x,y)\,dx\,dy, \tag{1.1}$$

where

$$H(x,y) = \mathbb{P}\{X \leqslant x, Y \leqslant y\} - \mathbb{P}\{X \leqslant x\}\,\mathbb{P}\{Y \leqslant y\}, \quad x, y \in \mathbb{R}. \tag{1.2}$$

The case of the identical functions $u(x) = x$ and $v(y) = y$ corresponds to Höffding [1] (provided that X and Y have finite second moments). The history of this remarkable identity can be found in [2], together with generalizations and refinements of previous results by Mardia [3], Sen [4], Cuadras [5, 6] (see also recent works [7, 8]).

* To whom the correspondence should be addressed.

Translated from *Problemy Matematicheskogo Analiza* **128**, 2024, pp. 19-32.

Block and Fang [9] proposed an extension of the original Höffding formula to more than two variables. Let us however restrict ourselves to the particular case of (1.1) with $X = Y$ and write this relation as a covariance identity with respect to the distribution μ of X:

$$\mathrm{cov}_\mu(u,v) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} u'(x)v'(y)\,d\lambda(x,y). \tag{1.3}$$

Here, one can require that λ be a positive, locally finite measure on the plane $\mathbb{R}\times\mathbb{R}$ (i.e., finite on compact sets). According to (1.1) and (1.2), the identity (1.3) holds if λ is absolutely continuous with respect to the Lebesgue measure with density

$$H_\mu(x,y) = \frac{d\lambda(x,y)}{dx\,dy} = F(x\wedge y)\,(1 - F(x\vee y)), \quad x,y\in\mathbb{R}, \tag{1.4}$$

where $F(x) = \mathbb{P}\{X \leqslant x\} = \mu((-\infty,x])$ is the associated distribution function. We adopt the standard notation $x\wedge y = \min(x,y)$ and $x\vee y = \max(x,y)$.

Definition 1.1. We call $\lambda = \lambda_\mu$ the *Höffding measure* and its density $H = H_\mu$ the *Höffding kernel* associated to μ.

For example, if $\mu = p\delta_a + q\delta_b$ is the Bernoulli measure assigning the weights $p\in(0,1)$ and $q = 1-p$ to the points $a<b$, then $\lambda_\mu = pqU$ where U is the uniform distribution on the square $(a,b)\times(a,b)$.

Let us state this consequence of (1.1)–(1.2) once more in the next statement with emphasis on the uniqueness part in the representation (1.3). As the weakest requirement, one can consider the latter identity in the class C_b^∞ of all functions $u,v:\mathbb{R}\to\mathbb{R}$ having C^∞-smooth, compactly supported derivatives (in which case, u and v are bounded).

Theorem 1.1. *Given a probability measure μ on the real line,* (1.3) *holds for all $u,v\in C_b^\infty$ with a unique positive, locally finite measure $\lambda=\lambda_\mu$. Moreover,* (1.3) *extends to all locally absolutely continuous, complex-valued functions u and v such that*

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |u'(x)|\,|u'(y)|\,d\lambda(x,y) < \infty, \qquad \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |v'(x)|\,|v'(y)|\,d\lambda(x,y) < \infty, \tag{1.5}$$

where the derivatives u' and v' are understood in the Radon–Nikodym sense. The Höffding measure λ is finite if and only if μ has a finite second moment.

The condition (1.5) insures that the function $u'(x)v'(y)$ is integrable over λ and also implies that $u(X)$ and $v(X)$ have finite second moments. Hence both sides in (1.3) are well-defined and finite. For the sake of completeness we sketch a short proof of Theorem 1.1 and give a few remarks on the existing (slightly different) formulations of this theorem at the end (Section 10).

Let us note that, being applied with $u(x) = v(x) = x$, the identity (1.3) shows that the total mass of the Höffding measure is the variance

$$\lambda(\mathbb{R}\times\mathbb{R}) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} H(x,y)\,dx\,dy = \mathrm{Var}(X),$$

which can be finite or not. Once the measure λ is finite, it can also be described via its Fourier–Stieltjes transform in terms of the characteristic function of the random variable X

$$f(t) = \mathbb{E}\, e^{itX} = \int_{-\infty}^{\infty} e^{itx}\, d\mu(x), \quad t \in \mathbb{R}.$$

Namely, applying (1.3) to the exponential functions $u(x) = e^{itx}$ and $v(y) = e^{isy}$ $(t, s \in \mathbb{R})$, we obtain the explicit formula

$$\widehat{\lambda}(t,s) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{itx+isy}\, d\lambda(x,y) = \frac{f(t)f(s) - f(t+s)}{ts} \quad (t, s \neq 0). \tag{1.6}$$

In particular, this provides the uniqueness part in Theorem 1.1 under the moment assumption $\mathbb{E}X^2 < \infty$.

As a consequence, let us also note that the expression on the right-hand side of (1.6) represents a positive definite function in two real variables, as long as the characteristic function f is twice differentiable (which guarantees the second moment condition).

2 Positive Definiteness of Höffding Kernels

Let X be a random variable with a nondegenerate distribution μ, so that the associated Höffding measure λ is nonzero. From (1.3) we have the variance representation

$$\mathrm{Var}(u(X)) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} u'(x)u'(y)\, d\lambda(x,y). \tag{2.1}$$

Hence the substitution $f = u'$ leads to the property

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x)f(y)H(x,y)\, dx\, dy \geqslant 0, \tag{2.2}$$

which holds for any measurable function f on the real line such that the last integral is well-defined in the Lebesgue sense. In other words, the following assertion holds.

Corollary 2.1. *Every Höffding kernel is positive definite: For any collection* $a_i, x_i \in \mathbb{R}$

$$\sum_{i,j=1}^{n} a_i a_j H(x_i, x_j) \geqslant 0. \tag{2.3}$$

Usually, the equivalence of (2.2) and (2.3) is stated under the assumption that a kernel is continuous. In the case of the Höffding kernels, this is however not important. Indeed, for the uniform distribution U on the unit interval $(0,1)$ the corresponding kernel $H_U(x,y) = (x \wedge y)\,(1 - x \vee y)$ is positive definite on the square $0 \leqslant x, y \leqslant 1$. Hence the same is true for $H_\mu(x,y) = H_U(F(x), F(y))$ on the plane.

Being positive definite, every Höffding kernel satisfies

$$H(x,y)^2 \leqslant H(x,x)H(y,y), \quad x, y \in \mathbb{R}, \tag{2.4}$$

which can be used to construct a pseudometric

$$d(x,y) = \big(H(x,x) - 2H(x,y) + H(y,y)\big)^{1/2}.$$

This property can be strengthened in terms of the Höffding measure λ. Since, by the Cauchy–Bunyakovsky–Schwarz inequality,

$$\operatorname{cov}_\mu(u,v)^2 \leqslant \operatorname{Var}_\mu(u)\operatorname{Var}_\mu(v),$$

we get from Theorem 1.1 that

$$\bigg(\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x)g(y)\,d\lambda(x,y)\bigg)^2 \leqslant \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x)f(y)\,d\lambda(x,y)\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x)g(y)\,d\lambda(x,y)$$

for all nonnegative measurable functions f and g on the real line. In particular,

$$\lambda(A\times B)^2 \leqslant \lambda(A\times A)\,\lambda(B\times B) \tag{2.5}$$

for all Borel sets $A, B \subset \mathbb{R}$. Hence (2.4) appears as an infinitesimal version of (2.5).

3 Spectral Decompositions

Since $H = H_\mu$ is positive definite, one can follow the advanced Mercer theory on metric spaces and develop the canonical representation

$$H(x,y) = \sum_{n=1}^{\infty} \alpha_n f_n(x) f_n(y) \tag{3.1}$$

in terms of eigenfunctions and eigenvalues of the linear operator

$$Tf(x) = \int_{-\infty}^{\infty} H(x,y)f(y)\,dy.$$

Let

$$a_0 = \inf\{x\in\mathbb{R} : F(x) > 0\}, \quad a_1 = \sup\{x\in\mathbb{R} : F(x) < 1\}, \tag{3.2}$$

where F is the distribution function of a random variable X with distribution μ. Applying Theorem 2.4 from [10] in the setting of the Höffding kernels, we get the following assertion.

Corollary 3.1. *Suppose that μ has finite first absolute moment and F is continuous. There exists an orthonormal system of continuous functions f_n in $L^2(a_0,a_1)$ and a nonincreasing sequence $\alpha_n \geqslant 0$ such that* (3.1) *holds for all $x, y \in (a_0, a_1)$. This series is absolutely and uniformly convergent on finite proper subintervals of (a_0,a_1).*

The moment assumption on μ guarantees that the Höffding kernel is square integrable over the rectangle $(a_0,a_1)\times(a_0,a_1)$. Moreover, T is acting on the Hilbert space $L^2(a_0,a_1)$ as a compact and selfadjoint operator. It is a trace class so that

$$\operatorname{Tr}(T) = \int_{a_0}^{a_1} H(x,x)\,dx = \int_{-\infty}^{\infty} F(x)(1-F(x))\,dx = \sum_{n=1}^{\infty}\alpha_n,$$

where the series is convergent. The latter integral can also be recognized as $\frac{1}{2}\,\mathbb{E}\,|X - X'|$ with X' being an independent copy of X.

As a consequence of (3.1), the variance representation (2.1) can be expressed in the form

$$\mathrm{Var}(u(X)) = \sum_{n=1}^{\infty} \alpha_n \Big(\int_{a_0}^{a_1} u'(x) f_n(x)\,dx \Big)^2.$$

For example, for the uniform distribution μ on the interval $(0,1)$, we have

$$F(x) = x, \quad H(x,y) = (x \wedge y)\,(1 - (x \vee y)) \quad (0 \leqslant x, y \leqslant 1).$$

In this case, (3.1) holds for all $x, y \in (0,1)$ with $\alpha_n = 1/(n\pi)^2$ and $f_n(x) = \sqrt{2}\,\sin(n\pi x)$.

More generally, suppose that μ has a continuous positive density $p(x)$ in $a_0 < x < a_1$. As easy to see, the spectral equation $Tf = \alpha f$ is reduced to the Sturm–Liouville equation

$$\alpha\,\Big(\frac{f'}{p}\Big)' + f = 0.$$

When the interval $[a_0, a_1]$ is finite, and p is continuous and positive on it, we thus arrive at the regular Sturm–Liouville problem with boundary conditions $f(a_0) = f(a_1) = 0$ for which the spectral theory is well-developed as well.

4 Marginals of Höffding Measures

Since the kernel $H = H_\mu$ is symmetric about the diagonal $x = y$, the Höffding measure $\lambda = \lambda_\mu$ has equal marginals $\Lambda = \Lambda_\mu$ defined by

$$\Lambda(A) = \lambda(A \times \mathbb{R}) = \int_A \int_{-\infty}^{\infty} H(x,y)\,dx\,dy, \quad A \subset \mathbb{R} \text{ (Borel)}. \tag{4.1}$$

It is absolutely continuous with respect to the Lebesgue measure and is supported on the interval (a_0, a_1), finite or not, defined in (3.2). Concerning its density, let us emphasize the following two simple properties.

Proposition 4.1. *If X has finite first absolute moment, then the marginal Λ is finite and has density*

$$h(x) = \frac{d\Lambda(x)}{dx} = \int_x^{\infty} (y - a)\,dF(y), \quad a = \mathbb{E}X. \tag{4.2}$$

In particular, it is unimodal with mode at the point a, i.e., $h(x)$ is nondecreasing on the half-axis $x < a$ and is nonincreasing for $x > a$. Moreover, it is continuous at $x = a$ with

$$h(a) = \frac{1}{2}\,\mathbb{E}\,|X - a|. \tag{4.3}$$

If $\mathbb{E}\,|X| = \infty$, then the density of Λ is almost everywhere infinite on (a_0, a_1).

Proof. According to (4.1), the measure Λ has density

$$h(x) = \int_{-\infty}^{\infty} F(x \wedge y)\,(1 - F(x \vee y))\,dy = (1 - F(x)) \int_{-\infty}^{x} F(y)\,dy + F(x) \int_{x}^{\infty} (1 - F(y))\,dy. \quad (4.4)$$

If $\mathbb{E}\,|X| = \infty$, then at least one of the two last integrals must be infinite for all $a_0 < x < a_1$, which means that $h(x) = \infty$ almost everywhere on (a_0, a_1).

If $\mathbb{E}\,|X| < \infty$, both integrals in (4.4) are finite. In addition, $F(y)y \to 0$ as $y \to -\infty$ and $(1 - F(y))y \to 0$ as $y \to \infty$. Assuming without loss of generality that F is continuous at the point x, one can integrate by parts to get

$$\int_{-\infty}^{x} F(y)\,dy = \int_{-\infty}^{x} F(y)\,d(y-a) = F(x)\,(x-a) - \int_{-\infty}^{x} (y-a)\,dF(y)$$

and similarly

$$\int_{x}^{\infty} (1 - F(y))\,dy = -(1 - F(x))\,(x-a) + \int_{x}^{\infty} (y-a)\,dF(y).$$

Hence the right-hand side of (4.4) becomes

$$\begin{aligned} &-(1 - F(x)) \int_{-\infty}^{x} (y-a)\,dF(y) + F(x) \int_{x}^{\infty} (y-a)\,dF(y) \\ &= F(x) \int_{-\infty}^{\infty} (y-a)\,dF(y) - \int_{-\infty}^{x} (y-a)\,dF(y) = \int_{x}^{\infty} (y-a)\,dF(y), \end{aligned}$$

which is (4.2).

Finally, since the function $x \to x - a$ is vanishing at the point a, it follows that

$$h(a-) = h(a+) = \int_{a}^{\infty} (y-a)\,dF(y) = \mathbb{E}\,(X-a)^+ = \frac{1}{2}\,\mathbb{E}\,|X-a|,$$

thus implying (4.3). □

Proposition 4.2. *Under the assumption that X has finite first absolute moment, the marginal Λ is a multiple of μ if and only if μ is Gaussian.*

Proof. Without loss of generality we can assume that $\mathbb{E}X = 0$. Since $\Lambda(\mathbb{R}) = \lambda(\mathbb{R} \times \mathbb{R})$, the property $\Lambda = \sigma^2 \mu$ with some constant σ^2 implies that Λ and λ are finite and forces μ to have a finite second moment. In that case, it follows from (1.6) that the Fourier–Stieltjes transform of Λ is given by

$$\widehat{\Lambda}(t) = \widehat{\lambda}(t, 0) = \int_{-\infty}^{\infty} e^{itx}\,h(x)\,dx = -\frac{f'(t)}{t}, \quad t \in \mathbb{R}, \quad t \neq 0,$$

where f is the characteristic function of X. Hence $\Lambda = \sigma^2 \mu$ if and only if

$$f'(t) = -\sigma^2 t f(t) \quad \forall\, t \in \mathbb{R}.$$

But this is only possible when μ is the Gaussian measure with mean zero and variance σ^2. □

Let us conclude with a few remarks. Often, the marginals of Höffding's measures appear in the particular case of the covariance representation (1.3) with the function $v(x) = x$. Then we arrive at

$$\operatorname{cov}(X, u(X)) = \int_{-\infty}^{\infty} u'(x)\, h(x)\, dx,$$

which holds as long as the integral is convergent. If μ is supported on an interval Δ and has there an almost everywhere positive density p, this formula can be written as

$$\operatorname{cov}(X, u(X)) = \mathbb{E}\, \tau(X) u'(X). \tag{4.5}$$

Here, the function

$$\tau(x) = \frac{h(x)}{p(x)} = \frac{1}{p(x)} \int_x^{\infty} (y - a)\, p(y)\, dy, \quad x \in \Delta,$$

is called the *Stein kernel*. We have $\tau(x) = 1$ (almost everywhere on Δ) if and only if μ is the standard Gaussian measure γ, in which case (4.5) becomes the Stein equation $\mathbb{E}\, X u(X) = \mathbb{E}\, u'(X)$.

After the pioneering work [11], the identity (4.5) served as a starting point in the extensive development of the Stein method as an approach to various forms of the central limit theorem and estimating the distances to the normal law γ avoiding the method of characteristic functions. For example, assuming that $\mathbb{E}X = 0$ and $\operatorname{Var}(X) = \sigma^2$, Cacoullos, Papathanasiou, and Utev [12] proposed a general upper bound for the total variation distance

$$\|\mu - \gamma\|_{\mathrm{TV}} \leqslant 4\, \mathbb{E}\, |\tau(X) - 1| + 4\, |1 - \sigma^2|$$

and applied it in the proof of the CLT with respect to this strong distance. For a comprehensive exposition of the whole theory, we refer an interested reader to the book [13].

5 Periodic Covariance Representations

A natural multidimensional extension of the covariance representation (1.3) for a given probability measure μ on $\mathbb{R}^n$ could be the identity

$$\operatorname{cov}_\mu(u, v) = \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} \langle \nabla u(x), \nabla v(y) \rangle \, d\lambda(x, y), \tag{5.1}$$

where λ is a suitable measure on $\mathbb{R}^n \times \mathbb{R}^n$. This is indeed possible when μ is Gaussian with covariance matrix $\sigma^2 I$ (a multiple of the identity matrix). In this case, λ is unique and can be described in several equivalent ways including Ornstein–Uhlenbeck semigroups and interpolation ([14, 15]). In [16], it was shown that the existence of λ in (5.1) forces the measure μ be Gaussian which gives another characterization of this class in terms of covariance representations. Nevertheless, some other variants of (5.1) could be applicable in order to involve larger classes of probability distributions in such identities. In particular, one can show that there is a spherical counterpart of (5.1)

$$\operatorname{cov}_{\sigma_{n-1}}(u, v) = \int_{S^{n-1}} \int_{S^{n-1}} \langle \nabla_S u(x), \nabla_S v(y) \rangle \, d\lambda(x, y) \tag{5.2}$$

with respect to the uniform distribution σ_{n-1} on the unit sphere S^{n-1} in $\mathbb{R}^n$ for some specific measure λ on $S^{n-1} \times S^{n-1}$ having multiples of σ_{n-1} as marginals. Here, $\nabla_S u$ denotes the spherical gradient of a smooth function u on the sphere.

However, the measures λ in such representations are not unique anymore. This can be seen already in the case of the circle S^1, when (5.2) is reduced to the covariance representation

$$\mathrm{cov}_\mu(u,v) = \int_{[0,1)} \int_{[0,1)} u'(x)v'(y)\, d\lambda(x,y) \tag{5.3}$$

for the uniform distribution $\mu = m$ on $(0,1)$ in the class of all 1-periodic smooth functions u and v on the real line. Here, λ is a finite measure on the square which we allow to be a signed measure for the sake of generality. Without loss of generality we require that it is symmetric about the diagonal $x = y$.

Keeping aside the multidimensional setting for a separate consideration, in what follows we focus on (5.3), assuming that μ is a given probability measure on $[0,1)$.

Definition 5.1. We call a signed symmetric Borel measure λ on $[0,1) \times [0,1)$ a *mixing measure* for μ if (5.3) holds for all 1-periodic smooth functions u and v on the real line.

As we discussed before, this identity always holds with the Höffding measure $\lambda = \lambda_\mu$. But its marginals can be a multiple of μ in the Gaussian case only. This motivates the following

Question. Given μ, is it possible to choose a mixing measure λ whose marginals are multiples of μ? If so, how can one describe all of them and choose a best one (in some sense)?

Towards this question, we prove the following assertion.

Theorem 5.1. *Let μ be a probability measure on $[0,1)$ with Höffding's measure λ_μ. Subject to the constraint that the marginal distribution of λ in (5.3) is equal to $c\mu$ for a prescribed value $c \in \mathbb{R}$, the mixing measure λ exists, is unique, and is given by*

$$\lambda = \lambda_\mu + (\sigma^2 - c)\, m \otimes m + c\,(\mu \otimes m + m \otimes \mu) - (\Lambda_\mu \otimes m + m \otimes \Lambda_\mu), \tag{5.4}$$

where m is the uniform distribution on $(0,1)$, Λ_μ is the marginal of the Höffding measure λ_μ associated to μ, and σ^2 is the variance of μ.

6 Covariance Representations for Uniform Distribution

We postpone the proof of Theorem 5.1 to Sections 8 and 9. The most interesting case in the periodic representation (5.3) is the one where μ represents a uniform distribution m on (0,1). Let us specialize Theorem 5.1 to this case and consider identities of the form

$$\mathrm{cov}_m(u,v) = \int_0^1 \int_0^1 u'(x)v'(y)\, d\lambda(x,y). \tag{6.1}$$

As a consequence of Theorem 5.1, we obtain the following statement needed in the study of covariance identities on the circle (which is a partial case of multidimensional spherical identities (5.2)).

Corollary 6.1. *Subject to the constraint that the marginal distribution of a mixing measure λ in* (6.1) *is equal to cm, $c \in \mathbb{R}$, the measure λ is unique and has density*

$$\frac{\lambda(x,y)}{dx\,dy} = D(|x-y|) + \Big(c - \frac{1}{24}\Big), \quad x, y \in (0,1), \tag{6.2}$$

where

$$D(h) = \frac{1}{8}\big[1 - 4h(1-h)\big], \quad 0 \leqslant h \leqslant 1. \tag{6.3}$$

Note that $D(h) \geqslant 0$ for all $h \in [0,1]$ and $D(h) = 0$ for $h = \frac{1}{2}$.

The mixing measure λ with density (6.2) is nonnegative if and only if $c \geqslant \frac{1}{24}$. Hence the smallest positive measure (for the usual comparison) corresponds to the parameter $c = \frac{1}{24}$, when λ has density $\psi(x,y) = D(|x-y|)$. In this sense, the optimal variant of (6.1) is given by the covariance representation

$$\mathrm{cov}_m(u,v) = \int_0^1\!\!\int_0^1 u'(x)v'(y)\,D(|x-y|)\,dx\,dy = \frac{1}{24}\int_0^1\!\!\int_0^1 u'(x)v'(y)\,d\nu(x,y)$$

with the probability measure $d\nu(x,y) = 24\,D(|x-y|)\,dx\,dy$ on $(0,1)\times(0,1)$. It has the uniform distribution m on $(0,1)$ as a marginal one.

Proof of Corollary 6.1. Returning to (5.4), note that, if μ has density p, then the measure λ has density

$$\psi(x,y) = H(x,y) + (\sigma^2 - c) + c\,(p(x) + p(y)) - (h(x) + h(y)) \tag{6.4}$$

on $[0,1)\times[0,1)$, where h denotes the density of the marginal Λ_μ of the Höffding measure λ_μ with density $H(x,y) = F(x\wedge y)\,(1 - F(x\vee y))$. Recall that according to (4.2),

$$h(x) = \int_x^1 yp(y)\,dy - a(1 - F(x)), \quad 0 \leqslant x \leqslant 1,$$

where a is the mean of μ.

In the case of the uniform distribution $\mu = m$, its distribution function and density are given by $F(x) = x$ and $p(x) = 1$ for $0 \leqslant x \leqslant 1$. Then, by (6.4),

$$\psi(x,y) = H(x,y) + (\sigma^2 + c) - (h(x) + h(y)) \tag{6.5}$$

with $H(x,y) = (x\wedge y)\,(1 - (x\vee y))$,

$$h(x) = \int_x^1 y\,dy - \frac{1}{2}\,(1-x) = \frac{1}{2}\,x(1-x),$$

and $\sigma^2 = \frac{1}{12}$. To simplify, assume that $0 \leqslant x \leqslant y \leqslant 1$. Then

$$\begin{aligned}\psi(x,y) &= c + \frac{1}{12} + x(1-y) - \frac{1}{2}\,((x - x^2) + (y - y^2))\\ &= c + \frac{1}{12} - \frac{1}{2}\,((y-x)(1-(y-x))) = c - \frac{1}{24} + D(y-x),\end{aligned}$$

thus proving (6.2). □

Remark 6.1. If we want to write down a similar representation on the interval $(0,T)$, $T>0$, we can use a linear transform. Let m_T denote the uniform distribution on $(0,T)$. Then we get that for all smooth T-periodic functions u and v and all $c \geqslant 1/24$

$$\mathrm{cov}_{m_T}(u,v) = \int_0^T \int_0^T u'(x)v'(y)\, d\lambda_T(x,y)$$

with a positive measure having the density

$$\frac{d\lambda_T(x,y)}{dx\, dy} = D\Big(\Big|\frac{x}{T} - \frac{y}{T}\Big|\Big) + \Big(c - \frac{1}{24}\Big)$$

on $(0,T)\times(0,T)$. It has the marginal $cTm_T(dx) = c\,dx$ on $(0,T)$.

7 Densities Bounded away from Zero

In the general situation, the question of whether or not the mixing measure λ is positive for a certain value of c is rather interesting (in which case this constant has to be positive as well). Here, we give one sufficient condition generalizing the previous example of the uniform distribution. As usual, we denote by σ the standard deviation of a random variable X distributed according to μ.

Corollary 7.1. *Suppose that a probability measure μ on $(0,1)$ has a density p such that $p(x) \geqslant \alpha$ for all $x \in (0,1)$ with some constant $\alpha > \frac{1}{2}$. There exists a positive mixing measure λ in the periodic covariance representation*

$$\mathrm{cov}_{\mu}(u,v) = \int_0^1 \int_0^1 u'(x)v'(y)\, d\lambda(x,y), \tag{7.1}$$

whose marginal is a multiple $c\mu$ of μ. One can choose

$$c = \frac{\sigma(1-\sigma)}{2\alpha - 1}.$$

Proof. According to (6.4), subject to the constraint that the marginal distribution of λ in (7.1) is equal to $c\mu$, the mixing measure λ has density

$$\psi(x,y) = H(x,y) + \sigma^2 + c\,(p(x)+p(y)-1) - (h(x)+h(y)), \quad x,y \in (0,1),$$

where $H(x,y)$ is the Höffding kernel and $h(x)$ is the density of the mariginal distribution Λ. Hence it is nonnegative as long as

$$c\,(p(x)+p(y)-1) \geqslant h(x)+h(y) - \sigma^2.$$

By assumption, $p(x)+p(y)-1 \geqslant 2\alpha - 1$, so that it suffices to require that

$$c\,(2\alpha - 1) \geqslant h(x)+h(y) - \sigma^2. \tag{7.2}$$

Now, let us recall that, by Proposition (4.1), $h(x)$ is unimodal and continuous. Moreover, according to (4.3), for all $x \in (0,1)$

$$2h(x) \leqslant 2h(a) = \mathbb{E}\,|X-a| \leqslant \sigma, \quad a = \mathbb{E}X,$$

where we applied the Cauchy inequality. Note that $\sigma^2 < \sigma$. Hence the right-hand side of (7.2) is bounded from above by $\sigma - \sigma^2$. □

Example 7.1. The symmetric beta distribution with parameters $(\frac{1}{2}, \frac{1}{2})$, i.e., with density

$$p(x) = \frac{1}{\pi\sqrt{x(1-x)}}, \quad 0 < x < 1,$$

satisfies the assumptions of Corollary 7.1 with $\alpha = 2/\pi$ and $\sigma^2 = 1/8$. Hence the conclusion in this corollary is true with

$$c = \frac{\pi}{8(4-\pi)}\,(\sqrt{8}-1) \sim 0.8364\ldots.$$

8 Characterization of Mixing Measures on Square

Let us first comment on the uniqueness issue in the problem of characterization of mixing measures λ on $[0,1) \times [0,1)$ in the periodic covariance representation

$$\mathrm{cov}_\mu(u,v) = \int\limits_{[0,1)} \int\limits_{[0,1)} u'(x)v'(y)\,d\lambda(x,y) \tag{8.1}$$

for a given probability measure μ on $[0,1)$. Applying this identity to the exponential functions $u(x) = e^{2\pi ikx}$ and $v(y) = e^{2\pi ily}$, we get the relation

$$\widehat{\mu}(k+l) - \widehat{\mu}(k)\,\widehat{\mu}(l) = -(2\pi)^2\,kl\,\widehat{\lambda}(k,l) \tag{8.2}$$

for all integers k, l, where

$$\widehat{\mu}(k) = \int\limits_0^1 e^{2\pi ikx}\,d\mu(x), \quad k \in \mathbb{Z},$$

and

$$\widehat{\lambda}(k,l) = \int\limits_0^1 \int\limits_0^1 e^{2\pi i(kx+ly)}\,d\lambda(x,y), \quad k,l \in \mathbb{Z},$$

denote the Fourier transforms of μ and λ restricted to integers. By the Stone–Weierstrass theorem applied on the circle, $\widehat{\lambda}$ determines any signed Borel measure λ on $[0,1) \times [0,1)$ in a unique way. This transform is explicitly defined in (8.2) as long as $k, l \neq 0$. Otherwise, both sides of (8.2) are vanishing. The fact that (8.2) does not define $\widehat{\lambda}$ for all integers does not allow us to reconstruct λ.

Moreover, due to the periodicity of u and v, we have

$$\int\limits_{[0,1)} \int\limits_{[0,1)} u'(x)v'(y)\,d\Lambda_1(x)\,dm(y) = \int\limits_{[0,1)} \int\limits_{[0,1)} u'(x)v'(y)\,dm(x)\,d\Lambda_2(y) = 0$$

for all signed measures Λ_1 and Λ_2 on $[0,1)$, where m denotes the uniform probability measure on that interval. Hence, once (8.1) is fulfilled for a measure λ, in particular, for the Höffding measure λ_μ, it is also fulfilled for

$$\lambda = \lambda_\mu + \Lambda_1 \otimes m + m \otimes \Lambda_2 \tag{8.3}$$

for any choice of signed measures Λ_1 and Λ_2 on $[0,1)$. We also have the converse statement (where the symmetry requirement is not required for a moment).

Lemma 8.1. *Let μ be a Borel probability measure on $[0,1)$. The covariance representation (8.1) holds for all C^1-smooth, 1-periodic functions u and v, if and only if λ has the form (8.3) for some (arbitrary) signed measures Λ_1 and Λ_2 on $[0,1)$.*

Proof. We only need to consider the necessity part. Assume that (8.1) holds for all C^1-smooth periodic functions u and v on the real line, so that

$$\int_{[0,1)} \int_{[0,1)} u'(x)v'(y)\, d\lambda_\mu(x,y) = \int_{[0,1)} \int_{[0,1)} u'(x)v'(y)\, d\lambda(x,y). \tag{8.4}$$

Putting $f = u'$, $g = v'$, we then have

$$\int_{[0,1)} \int_{[0,1)} f(x)g(y)\, d\lambda_\mu(x,y) = \int_{[0,1)} \int_{[0,1)} f(x)g(y)\, d\lambda(x,y). \tag{8.5}$$

Due to the periodicity of u and v, necessarily

$$\int_0^1 f(x)\, dm(x) = \int_0^1 g(y)\, dm(y) = 0 \tag{8.6}$$

and

$$f(0) = f(1), \quad g(0) = g(1). \tag{8.7}$$

Conversely, starting from continuous f an g on $[0,1]$ satisfying (8.6) and (8.7), we can define the functions

$$u(x) = \int_0^x f(t)\, dt, \quad v(y) = \int_0^y g(s)\, ds,$$

which have C^1-smooth 1-periodic extensions from $[0,1)$ to the whole real line and satisfy (8.4). Thus, the property (8.4) is equivalent to (8.5) subject to (8.6) and (8.7).

Let us reformulate the latter by identifying $[0,1)$ with the circle S^1 via the map $x \to e^{2\pi i x}$. It pushes forward m to the uniform probability measure σ_1 on the circle and pushes $\lambda - \lambda_\mu$ to some signed measure L on the torus $S^1 \times S^1$ with total mass $L(S^1 \times S^1) = 0$. Hence (8.5) subject to (8.6) and (8.7) is the same as the requirement

$$\int_{S^1} \int_{S^1} \xi(t)\eta(s)\, dL(t,s) = 0 \tag{8.8}$$

in the class of all continuous functions ξ, η on the circle such that

$$\int\limits_{S^1} \xi \, d\sigma_1 = \int\limits_{S^1} \eta \, d\sigma_1 = 0.$$

The latter assumption can be dropped if we write (8.8) as

$$\int\limits_{S^1} \int\limits_{S^1} (\xi(t) - \overline{\xi}) \, (\eta(s) - \overline{\eta}) \, dL(t,s) = 0, \tag{8.9}$$

where

$$\overline{\xi} = \int\limits_{S_1} \xi \, d\sigma_1, \quad \overline{\eta} = \int\limits_{S_1} \eta \, d\sigma_1.$$

At this step, by a simple approximation argument, (8.9) is extended to the class of all bounded, Borel measurable functions ξ and η on S^1.

Now, using the marginal measures

$$L_1(A) = L(A \times S^1), \quad L_2(B) = L(S^1 \times B) \quad (A, B \subset S^1 \text{ Borel sets}),$$

one can now write the equality (8.9) as

$$\int\limits_{S^1} \int\limits_{S^1} \xi(t)\eta(s) \, dL(t,s) - \int\limits_{S^1} \xi(t) \, d\sigma_1(t) \int\limits_{S^1} \eta(s) \, dL_2(s) - \int\limits_{S^1} \xi(t) \, dL_1(t) \int\limits_{S^1} \eta(s) \, d\sigma_1(s) = 0,$$

i.e.,

$$\int\limits_{S^1} \int\limits_{S^1} \xi(t)\eta(s) \, dK(t,s) = 0 \tag{8.10}$$

in terms of the measure $K = L - \sigma_1 \otimes L_2 - L_1 \otimes \sigma_1$. Moreover, (8.10) continues to hold for all finite linear combinations of functions of the form $\xi(t)\eta(s)$, $t, s \in S^1$. Therefore, by the approximation argument, this equality can be extended to all bounded, Borel measurable functions on the torus $S^1 \times S^1$. But this means that $K = 0$, which is an equivalent form of (8.3). $\square$

9 Proof of Theorem 5.1

Recall that the mixing measure λ in the periodic covariance representation (5.3) is supported on $[0,1) \times [0,1)$ and is required to be symmetric about the main diagonal of this square. This is fulfilled for the Höffding measure λ_μ. In general, by Lemma 8.1, λ must be of the form (8.3) for some signed measures Λ_1 and Λ_2 on $[0,1)$. But then λ is symmetric about the diagonal of the square if and only if $\Lambda_1 \otimes m + m \otimes \Lambda_2 = m \otimes \Lambda_1 + \Lambda_2 \otimes m$, i.e., $(\Lambda_1 - \Lambda_2) \otimes m = m \otimes (\Lambda_1 - \Lambda_2)$, where m is the uniform measure on $(0,1)$. This is equivalent to the statement that $\Lambda_1 - \Lambda_2$ is a multiple of m. In other words, the class of all symmetric measures λ satisfying the covariance representation (5.3) is described by the formula

$$\lambda = \lambda_\mu + b \, m \otimes m + \Lambda \otimes m + m \otimes \Lambda \tag{9.1}$$

with arbitrary $b \in \mathbb{R}$ and arbitrary signed measure Λ on $[0,1)$.

Such measures have equal marginals $\Lambda_\mu + (b+Q)\,m + \Lambda$, where $Q = \Lambda([0,1))$ and Λ_μ is the marginal of λ_μ described in (4.1). We want this measure to be a multiple of the original probability measure μ on $[0,1)$, i.e., $c\mu = \Lambda_\mu + (b+Q)\,m + \Lambda$ for some prescribed value $c \in \mathbb{R}$. Then necessarily with some $d \in \mathbb{R}$

$$\Lambda = c\mu - \Lambda_\mu + dm.$$

To determine the value of d, we insert this into (9.1) and get

$$\begin{aligned} \lambda &= \lambda_\mu + b\,m \otimes m + (c\mu - \Lambda_\mu + dm) \otimes m + m \otimes (c\mu - \Lambda_\mu + dm) \\ &= \lambda_\mu + (b+2d)\,m \otimes m + c\,(\mu \otimes m + m \otimes \mu) - (\Lambda_\mu \otimes m + m \otimes \Lambda_\mu). \end{aligned} \tag{9.2}$$

On marginals, this equality becomes the relation

$$b + 2d = \sigma^2 - c, \tag{9.3}$$

where $\sigma^2 = \Lambda_\mu\big([0,1)\big) = \lambda_\mu\big([0,1) \times [0,1)\big)$ is the variance of a random variable distributed according to μ. It remains to apply (9.3) in (9.2). Theorem 5.1 is now proved.

10 Proof of Theorem 1.1

One can assume that the locally absolutely functions u and v are real-valued and have Borel measurable Radon–Nikodym derivatives u' and v' (which are locally integrable).

Step 1. Assume that u' and v' are nonnegative. In view of the monotonicity of u and v, the covariance of $u(X)$ and $v(X)$ is well-defined and is given by

$$\begin{aligned} \mathrm{cov}\,(u(X), v(X)) &= \iint\limits_{t<s} (u(t) - u(s))(v(t) - v(s))\, d\mu(t)\, d\mu(s) \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} u'(x)v'(y)\, 1_{\{t \leqslant x < s\}}\, 1_{\{t \leqslant y < s\}}\, dx\, dy\ d\mu(t)\, d\mu(s). \end{aligned}$$

Since

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} 1_{\{t \leqslant x < s\}}\, 1_{\{t \leqslant y < s\}}\, d\mu(t)\, d\mu(s) = F(x \wedge y)\,(1 - F(x \vee y)),$$

an application of the Fubini theorem yields

$$\mathrm{cov}(u(X), v(X)) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} u'(x)v'(y)\, H(x,y)\, dx\, dy. \tag{10.1}$$

In particular,

$$\mathrm{Var}(u(X)) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} u'(x)u'(y)\, H(x,y)\, dx\, dy, \tag{10.2}$$

and similarly for v. As a byproduct, using $|\mathrm{cov}(u(X), v(X))|^2 \leqslant \mathrm{Var}(u(X))\ \mathrm{Var}(v(X))$, we have

$$\Big(\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} u'(x)v'(y)\, H(x,y)\, dx\, dy\Big)^2$$

$$\leqslant \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} u'(x)u'(y)\, H(x,y)\, dx\, dy \ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} v'(x)v'(y)\, H(x,y)\, dx\, dy. \tag{10.3}$$

Step 2. In the general case, define locally absolutely continuous, nondecreasing functions

$$\widetilde{u}(x) = \int_0^x |u'(t)|\, dt, \quad \widetilde{v}(x) = \int_0^x |v'(t)|\, dt,$$

which have Radon–Nikodym derivatives $|u'|$ and $|v'|$. Since $|\widetilde{u}(x) - \widetilde{u}(y)| \geqslant |u(x) - u(y)|$ for all $x, y \in \mathbb{R}$, we have

$$\mathrm{Var}(\widetilde{u}(X)) \geqslant \mathrm{Var}(u(X)),$$

and similarly for v. By Step 1,

$$\mathrm{Var}(\widetilde{u}(X)) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |u'(x)|\, |u'(y)|\, H(x,y)\, dx\, dy,$$

and the same is true for $\widetilde{v}$. Since this and a similar integral for v are supposed to be finite, we conclude that both $u(X)$ and $v(X)$ have finite second moments.

One can now repeat the arguments from Step 1, using the inequality (10.3) with $|u'|$ and $|v'|$ in place of u' and v' respectively. This will justify an application of the Fubini theorem, and then we obtain the identity (10.1) under the conditions in (1.5). This also insures the integrability of $u'(x)v'(y)$ over λ as a consequence of (1.5) and (10.2).

Step 3. For the uniqueness issue, let λ be a locally finite measure on the plane satisfying (1.3) in the class C_b^∞. Hence (after the replacement $u' = f$ and $v' = g$)

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x)g(y)\, d\lambda(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x)g(y)\, d\lambda_\mu(x,y)$$

for all C^∞-smooth, compactly supported functions $f, g : \mathbb{R} \to \mathbb{R}$. Using a suitable approximation, we conclude that $\lambda(A \times B) = \lambda_\mu(A \times B)$ for all bounded closed intervals A and B. Hence this equality is true for all Borel subsets of $\mathbb{R}^2$. The theorem is proved.

Remark 10.1. In [2, Theorem 3.1], the identity (1.1) is proved, assuming that u and v are absolutely continuous (not just locally), i.e.,

$$\int_{-\infty}^{\infty} |u'(x)|\, dx < \infty, \quad \int_{-\infty}^{\infty} |v'(x)|\, dx < \infty, \tag{10.4}$$

and such that the random variables $u(X)$, $v(Y)$, $u(X)v(Y)$ have finite first absolute moments. Note, however, that the condition (10.4) insures that both u and v are bounded, so that the moment assumptions are fulfilled automatically. A similar assertion with $X = Y$ is given in [7, Corollary 4], where in addition to the absolute continuity it is assumed that

$$\mathbb{E}\,|u(X)|^p < \infty, \quad \mathbb{E}\,|v(X)|^q < \infty \tag{10.5}$$

for some $p, q \geqslant 1$ such that $1/p + 1/q = 1$. Again, the latter assumption is not needed if we assume (10.4). As for the more general case of locally absolutely continuous u and v, the condition (10.5) and even the assumption on the boundedness of these functions do not guarantee that the integral in (1.3) is convergent in the Lebesgue sense, i.e.,

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |u'(x)|\,|v'(y)|\,H_\mu(x,y)\,dx\,dy < \infty.$$

For example, for $u(x) = v(x) = \cos x$, this integral is divergent as long as $\mathbb{E}\,|X| = \infty$.

Funding

Research of the first author was partially supported by BSF grant 2016050 and NSF grant DMS-2154001.

Declarations

Data availability This manuscript has no associated data.

Ethical conduct Not applicable.

Conflicts of interest The author declares that there is no conflict of interest.

References

1. W. Höffding, “Maszstabinvariante Korrelationstheorie,” *Schr. Math. Inst. u. Inst. Angew. Math. Univ. Berlin* **5**, 181–233 (1940);

2. A. Lo, “Functional generalizations of Hoeffding's covariance lemma and a formula for Kendall's tau,” *Stat. Probab. Lett.* **122** 218–226 (2017).

3. K. V. Mardia, “Some contributions to contingency-type bivariate distributions,” *Biometrika* **54**, 235–249 (1967).

4. P. K. Sen, “The impact of W. Hoeffding's research on nonparametrics,” In: *The Collected Works of Wassily Hoeffding*, Springer, New York (1994).

5. C. M. Cuadras, “On the covariance between functions,” *J. Multivariate Anal.* **81**, No. 1, 19–27 (2002).

6. C. M. Cuadras, “Contributions to the diagonal expansion of a bivariate copula with continuous extensions,” *J. Multivariate Anal.* **139**, 28–44 (2015).

7. A. Saumard and J. A. Wellner, “Efron's monotonicity property for measures on $\mathbb{R}^2$.,” *J. Multivariate Anal.* **166**, 212–224 (2018).

8. A. Saumard and J. A. Wellner, “On the isoperimetric constant, covariance inequalities and L_p-Poincaré inequalities in dimension one,” *Bernoulli* **25**, No. 3, 1794–1815 (2019).

9. H. W. Block and Z. B. Fang, “A multivariate extension of Hoeffding’s lemma,” *Ann. Probab.* **16**, No. 4, 1803–1820 (1988).

10. J. C. Ferreira and V. A. Menegatto, “Eigenvalues of integral operators defined by smooth positive definite kernels,” *Integral Equations Oper. Theory* **64**, No. 1, 61–81 (2009).

11. C. Stein, “A bound for the error in the normal approximation to the distribution of a sum of dependent random variables,” *Proc. 6th Berkeley Symposium on Math. Stat. and Probab.* **2**, 583–602 (1972).

12. T. Cacoullos, V. Papathanasiou, and S. A. Utev, “Variational inequalities with examples and an application to the central limit theorem,” *Ann. Probab.* **22**, No. 3, 1607–1618 (1994).

13. L. H. Y. Chen, L. Goldstein, and Q.-M. Shao, *Normal Approximation by Stein’s Method*, Springer, Berlin (2011).

14. C. Houdré and V. Pérez-Abreu, “Covariance identities and inequalities for functionals on Wiener and Poisson spaces,” *Ann. Probab.* **23**, No. 1, 400–419 (1995).

15. M. Ledoux, “L’algébre de Lie des gradients itérés d’un générateur Markovien – développements de moyennes et entropies,” *Ann. Sci. École. Norm. Sup.* **28**, 435–460 (1995).

16. S. G. Bobkov, F. Götze, and C. Houdré, “On Gaussian and Bernoulli covariance representations,” *Bernoulli* **7**, No. 3, 439–451 (2001).

Submitted on August 15, 2024