

HÖFFDING'S KERNELS AND PERIODIC COVARIANCE REPRESENTATIONS

SERGEY G. BOBKOV¹ AND DEVRAJ DUGGAL²

ABSTRACT. We start with a brief survey on Höfdding's kernels, its properties, related spectral decompositions, and discuss marginal distributions of Höfdding 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. Höfdding'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.

1. Generalized Höfdding's formula

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

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

where

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

The case of the identical functions $f(x) = x$ and $g(y) = y$ corresponds to Höfdding [8] (provided that X and Y finite second moments). The history of this remarkable identity may be found in Lo [11], together with generalizations and refinements of the previous results by Mardia [12], Sen [13], Cuadras [5, 6]; see also recent works by Saumard and Wellner [14, 15].

Block and Fang [1] proposed an extension of the original Höfdding's formula to more than two variables. Let us however restrict ourselves to the particular case of (1.1) with $X = Y$ and rewrite this relation as a covariance identity with respect to the distribution μ of X :

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

Here, one may require that λ be a positive, locally finite measure on the plane $\mathbb{R} \times \mathbb{R}$ (that is, finite on compact sets). According to (1.1)-(1.2), the identity (1.3) holds true, when λ is absolutely continuous over the Lebesgue measure with density

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

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

2010 *Mathematics Subject Classification*. Primary 60E, 60F.

Key words and phrases. Covariance representations, Höfdding's kernels.

1) School of Mathematics, University of Minnesota, Minneapolis, MN, USA .

2) School of Mathematics, University of Minnesota, Minneapolis, MN, USA .

Definition. 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 may consider the latter identity in the class C_b^∞ of all functions $u, v : \mathbb{R} \rightarrow \mathbb{R}$ 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 true 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, v such that*

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

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 in 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

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

Once this measure is finite, it may 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 an explicit formula

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

In particular, this provides the uniqueness part in Theorem 1.1.

Thus, 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. Moreover, there is a similar property of Höffding's kernels themselves.

2. Positive definiteness of Höfdding's kernels

Let X be a random variable with a non-degenerate distribution μ , so that the associated Höfdding measure λ is non-zero. From (1.5) we have a variance representation

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

and the substitution $f = u'$ leads to

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

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

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

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

Usually, the equivalence of (2.2) and (2.3) is stated under the assumption that a kernel is continuous. In the case of Höfdding's 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 $0 \leq x, y \leq 1$. Hence the same is true for $H_\mu(x, y) = H_U(F(x), F(y))$ on the plane.

Being positive definite, every Höfdding kernel satisfies

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

which may be used to construct a pseudometric

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

This property can be strengthened in terms of the Höfdding measure λ . Since, by the Cauchy inequality, $\text{cov}_\mu(u, v)^2 \leq \text{Var}_\mu(u)\text{Var}_\mu(v)$, we get from Theorem 1.1 that

$$\left(\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x)g(y) d\lambda(x, y) \right)^2 \leq \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x)f(y) d\lambda(x, y) \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x)g(y) d\lambda(x, y)$$

for all non-negative measurable functions f and g on the real line. In particular,

$$\lambda(A \times B)^2 \leq \lambda(A \times A) \lambda(B \times B) \quad (2.5)$$

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 may follow the advanced Mercer's theory on metric spaces and develop a canonical representation

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

in terms of the eigenfunctions and eigenvalues of the linear operator

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

Let

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

where F is the distribution function of X . Applying Theorem 2.4 from Ferreira and Menegatto [7] in the setting of Höfdding's kernels, we obtain:

Corollary 3.1. *Suppose that μ has finite first absolute moment, and F is continuous on the interval (a_0, a_1) . There exists an orthonormal system of continuous functions f_n in $L^2(a_0, a_1)$ and a non-increasing sequence $\alpha_n \geq 0$ 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 Höfdding's kernel is square integrable over the rectangle $(a_0, a_1) \times (a_0, a_1)$. In particular, T is acting on the Hilbert space $L^2(a_0, a_1)$ as a compact and self-adjoint operator. Moreover, it is a trace class so that

$$\text{tr}(T) = \int_a^b H(x, x) dx = \int_{-\infty}^{\infty} F(x)(1 - F(x)) dx = \sum_{n=1}^{\infty} \alpha_n,$$

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

As a consequence of (2.6), the variance representation (2.1) may be expressed in the form

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

For example, for the uniform distribution $\mu = U$, (2.6) holds for all $x, y \in (0, 1)$ with $\alpha_n = 1/(n\pi)^2$ and $f_n(x) = \sqrt{2} \sin(n\pi x)$.

In the general situation, let μ have 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

$$\alpha \left(\frac{f'}{p} \right)' + f = 0.$$

When p is continuous and positive on the finite interval $[a_0, a_1]$, 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 Höfdding measures

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

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

Obviously, 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*

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

In particular, it is unimodal with mode at the point a , that is, $h(x)$ is non-decreasing on the half-axis $x < a$ and is non-increasing for $x > a$. Moreover, it is continuous at $x = a$ with

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

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

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

$$\begin{aligned} h(x) &= \int_{-\infty}^\infty F(x \wedge y) (1 - F(x \vee y)) dy \\ &= (1 - F(x)) \int_{-\infty}^x F(y) dy + F(x) \int_x^\infty (1 - F(y)) dy. \end{aligned} \quad (4.4)$$

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 $h(x) = \infty$ a.e. on (a_0, a_1) .

If $\mathbb{E} |X| < \infty$, both integrals in (4.4) are finite. Assuming without loss of generality that F is continuous at the point x , one may integrate by parts in (4.4) to obtain (4.2). Finally, since the function $x \rightarrow x - a$ is vanishing at the point a , it follows that

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

A similar equality is also true for $h(a-)$, thus implying (4.3). \square

Proposition 4.2. *Assuming 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 may 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

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

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

$$f'(t) = -\sigma^2 t f(t) \quad \text{for all } t \in \mathbb{R}.$$

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

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

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

holding true as long as the integral is convergent. If μ is supported on an interval Δ and has there an a.e. positive density p , this formula may be rewritten as

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

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 $\tau(x) = 1$ (a.e. on Δ), if and only if μ is the standard Gaussian measure γ , in which case (4.5) becomes Stein's equation $\mathbb{E} X u(X) = \mathbb{E} u'(X)$. After the pioneering work [17], the identity (4.5) served as a starting point in the extensive development of Stein's 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, Cacoullos, Papathansiou and Utev [3] proposed a general upper bound for the total variation distance

$$\|\mu - \gamma\|_{\text{TV}} \leq 4 \mathbb{E} |\tau(X) - 1| + 4|1 - a|$$

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 by Chen, Goldstein and Shao [4] and survey [16].

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

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

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 ([10], [9]). In [2], 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),

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

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

$$\text{cov}_\mu(u, v) = \int_0^1 \int_0^1 u'(x)v'(y) d\lambda(x, y) \quad (5.3)$$

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

As we discussed before, this identity always holds with the Höfding measure $\lambda = \lambda_\mu$. But, its marginals may 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 to describe all of them and choose a best one (in some sense)?

Towards this question, we prove the next assertion.

Theorem 5.1. *Let μ be a probability measure on $[0, 1)$ with the Höfding 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*

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

where Λ_μ is the marginal of λ_μ and σ^2 is the variance of μ .

6. Covariance representations for the uniform distribution

We postpone the proof of Theorem 5.1 to Sections 8-9. 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

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

As a consequence of Theorem 5.1, we obtain the following statement needed in the study of covariance identities on the circle (as part 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*

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

where

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

Note that $D(h) \geq 0$ for all $h \in [0, 1]$ and the inequality becomes an equality for $h = \frac{1}{2}$.

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

$$\begin{aligned} \text{cov}_m(u, v) &= \int_0^1 \int_0^1 u'(x)v'(y) D(|x - y|) dx dy \\ &= \frac{1}{24} \int_0^1 \int_0^1 u'(x)v'(y) d\nu(x, y) \end{aligned}$$

with a 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. Returning to (5.4), note that, if μ has density p , then the measure λ has density

$$\psi(x, y) = h(x, y) + (\sigma^2 - c) + c(p(x) + p(y)) - (q(x) + q(y)) \quad (6.4)$$

on $[0, 1) \times [0, 1)$, where h denotes the density of the marginal Λ_μ of the Höfdding measure λ_μ with density $H(x, y) = F(x \wedge y) (1 - F(x \vee y))$. Recall that according to (4.1), if the mean of μ is $a = \mathbb{E}X$, then

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

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

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

with $H(x, y) = (x \wedge y) (1 - (x \vee y))$ and $\sigma^2 = \frac{1}{12}$. To simplify, assume that $0 \leq x \leq y \leq 1$. Then

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

□

Remark. If we want to write down a similar representation on the interval $(0, T)$, $T > 0$, one may 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 for all $c \geq 1/24$,

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

with a positive measure having the density

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

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

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

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

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

Proof. According to (5.4), subject to the constraint that the marginal distribution of λ in (7.1) is equal to $c\mu$, the mixing measure λ has density

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

where $H(x, y)$ is the Höfdding kernel and $h(x)$ is the density of the mariginal distribution Λ . Hence, it is non-negative, as long as

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

By the assumption, $p(x) + p(y) - 1 \geq 2\alpha - 1$, so that, it is sufficient to require that

$$c(2\alpha - 1) \geq h(x) + h(y) - \sigma^2. \quad (7.2)$$

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

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

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

Example. The symmetric beta distribution with parameters $(\frac{1}{2}, \frac{1}{2})$, that is, with density

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

satisfies the conditions of Corollary 7.1 with $\alpha = \frac{2}{\pi}$ and $\sigma^2 = \frac{1}{8}$. Hence, the conclusion in this corollary is true with $c = \frac{\pi}{8(4-\pi)}(\sqrt{8}-1) \sim 0.8364\dots$. In fact, a more careful analysis shows that one may take $c = \frac{1}{8}$ in this example (which is optimal).

8. Characterization of mixing measures

Let us first comment on the uniqueness issue in the problem of characterization of mixing measures in the covariance representation (5.3). Applying this identity to the exponential functions $u(x) = e^{2\pi i k x}$ and $v(y) = e^{2\pi i l y}$, we get the relation

$$f(k+l) - f(k)f(l) = -(2\pi)^2 kl \hat{\lambda}(k, l) \quad (8.1)$$

for all integers k, l , where

$$\hat{\lambda}(k, l) = \int_0^1 \int_0^1 e^{2\pi i(kx+ly)} d\lambda(x, y), \quad k, l \in \mathbb{Z},$$

denotes the Fourier transform of λ restricted to integers. By the Stone-Weierstrass theorem (applied on the circle), $\hat{\lambda}$ determines any signed Borel measure λ on $[0, 1) \times [0, 1)$ in a unique way. This transform is explicitly defined in (8.1) as long as $k, l \neq 0$. Otherwise, both sides of (8.1) are vanishing. The fact that (8.1) does not define $\hat{\lambda}$ for all integers does not allow us to reconstruct λ .

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

$$\int_0^1 \int_0^1 u'(x)v'(y) d\Lambda_1(x) dm(y) = \int_0^1 \int_0^1 u'(x)v'(y) dm(x) d\Lambda_2(y) = 0$$

for all signed measures λ_1 and λ_2 on $[0, 1)$, where m denotes the uniform probability measure on that interval. Hence, once (5.3) is fulfilled for a measure λ , in particular, for the Höfding measure λ_μ , it is also fulfilled for

$$\lambda = \lambda_\mu + \Lambda_1 \otimes m + m \otimes \Lambda_2 \quad (8.2)$$

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*

$$\text{cov}_\mu(u, v) = \int_0^1 \int_0^1 u'(x)v'(y) d\lambda(x, y) \quad (8.3)$$

holds true for all smooth, 1-periodic functions u and v , if and only if λ has the form (8.2) for some (arbitrary) signed measures Λ_1 and Λ_2 on $[0, 1)$.

Proof. We only need to consider the necessity part. Assume that (8.3) holds true for all C^1 -smooth periodic functions u and v , so that

$$\int_0^1 \int_0^1 u'(x)v'(y) d\lambda_\mu(x, y) = \int_0^1 \int_0^1 u'(x)v'(y) d\lambda(x, y). \quad (8.4)$$

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

$$\int_0^1 \int_0^1 f(x)g(y) d\lambda_\mu(x, y) = \int_0^1 \int_0^1 f(x)g(y) d\lambda(x, y). \quad (8.5)$$

Due to the periodicity of u and v , necessarily

$$\int_0^1 f(x) dm(x) = \int_0^1 g(y) dm(y) = 0 \quad (8.6)$$

and

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

Conversely, starting from continuous f and g on $[0, 1]$ satisfying (8.6)–(8.7), we may 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, our hypothesis (8.3) is equivalent to (8.5) subject to (8.6)–(8.7).

Let us reformulate the latter by identifying $[0, 1)$ with the circle S^1 via the map $x \rightarrow e^{2\pi ix}$. 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 $S^1 \times S^1$. Hence, (8.5) subject to (8.6)–(8.7) is the same as the requirement

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

in the class of all continuous functions ξ, η on the circle such that $\int_{S^1} \xi d\sigma_1 = \int_{S^1} \eta d\sigma_1 = 0$. The latter assumption may be dropped, if we rewrite (8.8) as

$$\int_{S^1} \int_{S^1} (\xi(t) - \bar{\xi})(\eta(s) - \bar{\eta}) dL(t, s) = 0, \quad (8.9)$$

where

$$\bar{\xi} = \int_{S^1} \xi d\sigma_1, \quad \bar{\eta} = \int_{S^1} \eta d\sigma_1.$$

In this step, (8.9) is readily extended to the class of all bounded, Borel measurable functions ξ and η on S^1 . Using the marginal measures

$$L_1(A) = L(A \times S^1), \quad L_2(B) = L(S^1 \times B),$$

one may now rewrite the equality (8.9) as

$$\begin{aligned} \iint \xi(t)\eta(s) dL(t, s) &= \int \xi(t) d\sigma_1(t) \int \eta(s) d\sigma_1(s) \\ &+ \int \xi(t) d\sigma_1(t) \int \eta(s) dL_2(s) + \int \xi(t) dL_1(t) \int \eta(s) d\sigma_1(s). \end{aligned}$$

But this means that

$$L = \sigma_1 \otimes \sigma_1 + \sigma_1 \otimes L_2 + L_1 \otimes \sigma_1.$$

Pushing σ_1 and L_j back to $[0, 1)$ with images Λ_j and L to $[0, 1) \times [0, 1)$, we arrive at

$$\lambda - \lambda_\mu = m \otimes m + m \otimes \Lambda_2 + \Lambda_1 \otimes m.$$

This is an equivalent form for (8.2). □

9. Proof of Theorem 5.1

Recall that the mixing measure λ in (5.3) has to be supported on $[0, 1) \times [0, 1)$ and is required to be symmetric about the diagonal of this square. This is fulfilled for the Höfding measure λ_μ . Hence, the measure λ of the form (8.2) is symmetric about the diagonal of the square, if and only if $\Lambda_2 - \Lambda_1$ is proportional to the uniform measure m . In other words, the class of all symmetric measures λ satisfying the covariance representation (5.3) is described by the formula

$$\lambda = \lambda_\mu + b m \otimes m + \Lambda \otimes m + m \otimes \Lambda \quad (9.1)$$

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

Such measures have equal marginals

$$\text{Proj}(\lambda) = \Lambda_\mu + (b + Q) m + \Lambda, \quad \Lambda_\mu = \text{Proj}(\lambda_\mu),$$

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 $[0, 1)$, that is,

$$c\mu = \Lambda_\mu + (b + Q) m + \Lambda$$

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 plug this into (9.1) and get

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

On marginals this equality becomes the relation

$$b + c + 2d - \sigma^2 = 0, \quad (9.3)$$

where

$$\sigma^2 = \Lambda_\mu([0, 1)) = \lambda_\mu([0, 1) \times [0, 1))$$

is variance of a random variable distributed according to μ . This is how d should be determined in terms of the free parameters b and c . In that case,

$$\Lambda = c\mu - \Lambda_\mu + \frac{\sigma^2 - b - c}{2} m. \quad (9.4)$$

It remains to apply (9.3) in (9.2). Theorem 5.1 is now proved.

10. Proof of Theorem 1.1

One may 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 have non-negative u' and v' . In view of the monotonicity of u and v , the covariance of $u(X)$ and $v(X)$ is well-defined and is given by

$$\begin{aligned} \text{cov}(u(X), v(X)) &= \iint_{x < y} (u(x) - u(y))(v(x) - v(y)) d\mu(x) d\mu(y) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u'(t)v'(s) 1_{\{x \leq t < y\}} 1_{\{x \leq s < y\}} dt ds d\mu(x) d\mu(y) \end{aligned}$$

as the Lebesgue integral on \mathbb{R}^4 over the product measure $L \otimes L \otimes \mu \otimes \mu$ (where L is the Lebesgue measure on the real line). By Fubini's theorem, we obtain that

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

In particular,

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

and similarly for v . As a by-product, using $|\text{cov}(u(X), v(X))|^2 \leq \text{Var}(u(X)) \text{Var}(v(X))$,

$$\left(\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u'(x)v'(y) H(x, y) dx dy \right)^2 \leq \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u'(x)u'(y) H(x, y) dx dy \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} v'(x)v'(y) H(x, y) dx dy. \quad (10.3)$$

Step 2. In the general case, define locally absolutely continuous, non-decreasing functions

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

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

$$\text{Var}(\tilde{u}(X)) \geq \text{Var}(u(X)),$$

and similarly for v . By the previous step,

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

and the same is true for \tilde{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 may 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's 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^∞ . Using a simple approximation, we obtain that $\lambda(A \times B) = \lambda_\mu(A \times B)$ for all bounded intervals A and B . Hence, this equality is true for all Borel subsets of \mathbb{R}^2 .

Remarks. In [11], Theorem 1.1 is proved in a more general setting of random variables X and Y as the identity (1.1), assuming that u and v are absolutely continuous (not just locally), that is,

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

and such that $u(X)$, $v(Y)$, $u(X)v(Y)$ have finite first absolute moments (cf. Theorem 3.1 in [11]). 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 [14], Corollary 4, where in addition to the absolute continuity it is assumed that

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

for some $p, q \geq 1$ such that $\frac{1}{p} + \frac{1}{q} = 1$. 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, that is,

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

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

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

REFERENCES

- [1] Block, H. W.; Fang, Z. B. A multivariate extension of Hoeffding's lemma. *Ann. Probab.* 16 (1988), no. 4, 1803–1820.
- [2] Bobkov, S. G.; Götze, F.; Houdré, C. On Gaussian and Bernoulli covariance representations. *Bernoulli* 7 (2001), no. 3, 439–451.
- [3] Cacoullos, T.; Papathansiou, V.; Utev, S. A. Variational inequalities with examples and an application to the central limit theorem. *Ann. Probab.* 22 (1994), 1607–1618.
- [4] Chen, L. H. Y.; Goldstein, L.; Shao, Q.-M. *Normal approximation by Stein's method. Probability and its Applications (New York).* Springer, Heidelberg, 2011. xii+405 pp.
- [5] Cuadras, C. M. On the covariance between functions. *J. Multivariate Anal.* 81 (2002), 19–27.
- [6] Cuadras, C. M. Contributions to the diagonal expansion of a bivariate copula with continuous extensions. *J. Multivariate Anal.* 139 (2015), 28–44.
- [7] Ferreira, J. C.; Menegatto, V. A. Eigenvalues of integral operators defined by smooth positive definite kernels. *Integral Equations Operator Theory* 64 (2009), no. 1, 61–81.
- [8] Hoeffding, K. Maszstabinvariante Korrelationstheorie. *Schr. Math. Inst. U. Inst. Angew. Math. Univ. Berlin* 5 (1940), 181–233.
- [9] Houdré, C.; Pérez-Abreu, V. Covariance identities and inequalities for functionals on Wiener and Poisson spaces. *Ann. Probab.* 23 (1995), 400–419.
- [10] Ledoux, M. L'algèbre de Lie des gradients itérés d'un générateur Markovien – développements de moyennes et entropies. *Ann. Sci. École. Norm. Sup.* 28 (1995), 435–460.
- [11] Lo, A. Functional generalizations of Hoeffding's covariance lemma and a formula for Kendall's tau. *Statistics and Probability Letters.* 122 (2017), 218–226.
- [12] Mardia, K. V. Some contributions to contingency-type bivariate distributions. *Biometrika* 54 (1967), 235–249.
- [13] Sen, P. K. The impact of Wassily Hoeffding's research on nonparametrics. In: *The Collected Works of Wassily Hoeffding.* Springer Science+Business Media, New York, 1994.
- [14] Saumard, A.; Wellner, J. A. Efron's monotonicity property for measures on \mathbb{R}^2 . *J. Multivariate Anal.* 166 (2018), 212–224.
- [15] Saumard, A.; Wellner, J. A. On the isoperimetric constant, covariance inequalities and L_p -Poincaré inequalities in dimension one. *Bernoulli* 25 (2019), no. 3, 1794–1815.
- [16] Shao, Q.-M. Stein's method, self-normalized limit theory and applications. *Proceedings of the International Congress of Mathematicians. Vol. IV, 2325–2350,* Hindustan Book Agency, New Delhi, 2010.
- [17] Stein, C. A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. In *Proc. of the Sixth Berkeley Symposium on Math. Stat. and Probab. (Univ. California, Berkeley, Calif., 1970/1971), Vol. II: Probability theory* 583–602, 1972, Berkeley, CA.