Hilbert space embeddings of independence tests in 2 and several variables

Jean Carlo Guella

University of Campinas (UNICAMP)-IMECC
Grants 2022/00008-0 and 2021/04226-0, São Paulo Research Foundation (FAPESP)

June 29, 2023

The objective of this talk is to understand what properties a function $\mathfrak{I} : [\prod_{i=1}^n X_i] \times [\prod_{i=1}^n X_i] \to \mathbb{R}$ that can be used as an independence test by an "integration method", that is

$$\int_{\prod_{i=1}^n X_i} \int_{\prod_{i=1}^n X_i} \mathfrak{I}(x,y) d[P - \otimes_{i=1}^n P_i](x) d[P - \otimes_{i=1}^n P_i](y) > 0$$

whenever $P \neq \otimes_{i=1}^n P_i$, must have.

The objective of this talk is to understand what properties a function $\mathfrak{I} : [\prod_{i=1}^{n} X_i] \times [\prod_{i=1}^{n} X_i] \to \mathbb{R}$ that can be used as an independence test by an "integration method", that is

$$\int_{\prod_{i=1}^{n} X_i} \int_{\prod_{i=1}^{n} X_i} \mathfrak{I}(x, y) d[P - \otimes_{i=1}^{n} P_i](x) d[P - \otimes_{i=1}^{n} P_i](y) > 0$$

whenever $P \neq \otimes_{i=1}^{n} P_i$, must have.
We will see that is convenient and in several scenarios equivalent to analyze this property on a larger set, more specifically when P and Q have the same marginals

$$\int_{\prod_{i=1}^{n} X_i} \int_{\prod_{i=1}^{n} X_i} \mathfrak{I}(x, y) d[P - Q](x) d[P - Q](y) > 0$$

whenever $P \neq Q$.

The objective of this talk is to understand what properties a function $\mathfrak{I} : [\prod_{i=1}^n X_i] \times [\prod_{i=1}^n X_i] \to \mathbb{R}$ that can be used as an independence test by an "integration method", that is

$$\int_{\prod_{i=1}^n X_i} \int_{\prod_{i=1}^n X_i} \mathfrak{I}(x,y) d[P - \otimes_{i=1}^n P_i](x) d[P - \otimes_{i=1}^n P_i](y) > 0$$

whenever $P \neq \otimes_{i=1}^n P_i$, must have.
We will see that is convenient and in several scenarios equivalent to analyze this property on a larger set, more specifically when P and Q have the same marginals

$$\int_{\prod_{i=1}^n X_i} \int_{\prod_{i=1}^n X_i} \mathfrak{I}(x,y) d[P - Q](x) d[P - Q](y) > 0$$

whenever $P \neq Q$.
One of the reasons is that by Hahn-Jordan decomposition the set $\{M[P-Q], M \geq 0, P, Q \text{ have the same marginals}\}$ is the vector space of finite measures such that $\mu(\otimes_{i=1}^n A_i) = 0$ whenever $|\{i \mid A_i = X_i\}| = n-1$.

Introduction: PD kernels

• A kernel $K : X \times X \to \mathbb{R}$ is called Positive Definite (PD) if it is symmetric and for whichever finite quantity of points $x_1, \dots, x_n \in X$ and scalars $c_1, \dots, c_n \in \mathbb{R}$

$$\sum_{i,j=1}^{n} c_i c_j K(x_i, x_j) \geq 0.$$

• A kernel $K : X \times X \to \mathbb{R}$ is called Positive Definite (PD) if it is symmetric and for whichever finite quantity of points $x_1, \dots, x_n \in X$ and scalars $c_1, \dots, c_n \in \mathbb{R}$

$$\sum_{i,j=1}^{n} c_i c_j K(x_i, x_j) \geq 0.$$

(**Kernel mean Embedding**) If K is continuous, $\mu \in \mathfrak{M}(X)$ is a finite Radon measure and $\sqrt{K(x,x)} \in L^1(|\mu|)$, then

$$y \to K_\mu(y) := \int_X K(x,y) d\mu(x) \text{ is an element of } \mathcal{H}_K$$

and if ν is another measure that satisfies the same conditions

$$\langle K_\mu, K_\nu \rangle_{\mathcal{H}_K} = \int_X \int_X K(x,y) d\mu(x) d\nu(y).$$

If K is bounded and $\mu \in \mathfrak{M}(X) \to \mathcal{H}_K$ is injective we say that K is **Integrally Strictly Positive Definite (ISPD)**. (In particular we obtain an inner product in $\mathfrak{M}(X)$)

[1] J. C. Guella, On Gaussian kernels on Hilbert spaces and kernels on Hyperbolic spaces, JAT (2022)

Introduction: PD kernels

If K is bounded and $\mu \in \mathfrak{M}(X) \to \mathcal{H}_K$ is injective we say that K is **Integrally Strictly Positive Definite (ISPD)**. (In particular we obtain an inner product in $\mathfrak{M}(X)$)
If K is bounded and injective in the subspace $\mathfrak{M}_0(X) := \{\mu \in \mathfrak{M}(X), \mu(X) = 0\}$, we say that K is **Characteristic**. (In particular we obtain an injective embedding of $\mathcal{P}(X)$ to a Hilbert space)

By the Hahn-Jordan decomposition, K is Characteristic if and only if

$$\int_X \int_X K(x,y)d[P-Q](x)d[P-Q](y) \geq 0, \quad P, Q \in \mathcal{P}(X)$$

and it is zero only when $P = Q$. Examples for ISPD include the Gaussian kernel in any Hilbert space and the majority in the Gneiting Class[1].

[1] J. C. Guella, On Gaussian kernels on Hilbert spaces and kernels on Hyperbolic spaces, JAT (2022)

• A kernel $\gamma : X \times X \to \mathbb{R}$ is called Conditionally Negative Definite (CND) if it is symmetric and for whichever finite quantity of points $x_1, \dots, x_n \in X$ and scalars $c_1, \dots, c_n \in \mathbb{R}$, restricted to $\sum_{i=1}^{n} c_i = 0$, we have that

$$\sum_{i,j=1}^{n} c_i c_j \gamma(x_i, x_j) \leq 0.$$

Introduction: CND kernels

• A kernel $\gamma : X \times X \to \mathbb{R}$ is called Conditionally Negative Definite (CND) if it is symmetric and for whichever finite quantity of points $x_1, \ldots, x_n \in X$ and scalars $c_1, \ldots, c_n \in \mathbb{R}$, restricted to $\sum_{i=1}^n c_i = 0$, we have that

$$\sum_{i,j=1}^{n} c_i c_j \gamma(x_i, x_j) \leq 0.$$

There is a strong connection between CND and PD kernels:

Theorem: A symmetric kernel $\gamma : X \times X \to \mathbb{R}$ is CND if and only if for any (or equivalently for every) $z \in X$ the kernel

$$K^\gamma(x, y) := \gamma(x, z) + \gamma(z, y) - \gamma(x, y) - \gamma(z, z) \quad \text{is PD, and}$$

$$2\gamma(x, y) - \gamma(x, x) - \gamma(y, y) = \|K_x^\gamma - K_y^\gamma\|_{\mathcal{H}_K}^2.$$

$$K^\gamma(x, y) = \int_X \int_X \gamma(u, v) d[\delta_x - \delta_z](u) d[\delta_y - \delta_z](v)$$

By the previous relation, if γ is continuous, bounded at the diagonal and $\mu \in \mathfrak{M}_0(X)$ is such that $\gamma \in L^1(|\mu| \times |\mu|)$ (the last relation is equivalent to $\gamma(\cdot, z) \in L^1(|\mu|)$ for every $z \in X$), then

$$\int_X \int_X -\gamma(x,y)d\mu(x)d\mu(y) = \int_X \int_X K^{\gamma}(x,y)d\mu(x)d\mu(y) \geq 0.$$

[2] J. C. Guella, Generalization of the energy distance by Bernstein functions, J Theo. Prob. (2022)

By the previous relation, if γ is continuous, bounded at the diagonal and $\mu \in \mathfrak{M}_0(X)$ is such that $\gamma \in L^1(|\mu| \times |\mu|)$ (the last relation is equivalent to $\gamma(\cdot, z) \in L^1(|\mu|)$ for every $z \in X$), then

$$\int_X \int_X -\gamma(x,y)d\mu(x)d\mu(y) = \int_X \int_X K^{\gamma}(x,y)d\mu(x)d\mu(y) \geq 0.$$

Similar to the definition of Characteristic kernels, if the previous inequality is zero only when μ is the zero measure we say that γ is CND-Characteristic (also "Strong negative type").

Examples include the Brownian kernel in any Hilbert space and the metric in real/complex hyperbolic spaces of any dimension. [2]

[2] J. C. Guella, Generalization of the energy distance by Bernstein functions, J Theo. Prob. (2022)

Since

$$-t^{1/2} = \frac{1}{2\sqrt{\pi}} \int_{(0,\infty)} (e^{-rt} - 1) \frac{dr}{r^{3/2}}, \quad t \geq 0$$

we can prove that $\|x - y\|$ is CND.

Since

$$-t^{1/2} = \frac{1}{2\sqrt{\pi}} \int_{(0,\infty)} (e^{-rt} - 1) \frac{dr}{r^{3/2}}, \quad t \geq 0$$

we can prove that $\|x - y\|$ is CND. If $\mu \in \mathfrak{M}(\mathcal{H})$, $\mu(\mathcal{H}) = 0$ and $\| \cdot \| \in L^1(|\mu|)$

$$\begin{aligned}\int_{\mathcal{H}} \int_{\mathcal{H}} -\|x - y\| d\mu(x) d\mu(y) &= \int_{\mathcal{H}} \int_{\mathcal{H}} \left[\frac{1}{2\sqrt{\pi}} \int_{(0,\infty)} (e^{-r\|x-y\|^2} - 1) \frac{dr}{r^{3/2}} \right] d\mu(x) d\mu(y) \\ &= \frac{1}{2\sqrt{\pi}} \int_{(0,\infty)} \left[\int_{\mathcal{H}} \int_{\mathcal{H}} e^{-r\|x-y\|^2} d\mu(x) d\mu(y) \right] \frac{dr}{r^{3/2}} \geq 0.\end{aligned}$$

It is possible to use the concepts of ISPD and CND-Characteristic kernels in order to obtain independence tests. Let P be a probability in $\mathcal{H} \times \mathcal{H}'$ and its marginals P_1, P_2, then

$$\int_{\mathcal{H}\times\mathcal{H}'}\int_{\mathcal{H}\times\mathcal{H}'} e^{-\|x-y\|^2}e^{-\|z-w\|^2} d[P - P_1 \otimes P_2](x,z)d[P - P_1 \otimes P_2](y,w) = 0$$

$$\int_{\mathcal{H}\times\mathcal{H}'}\int_{\mathcal{H}\times\mathcal{H}'} \|x-y\|\|z-w\| d[P - P_1 \otimes P_2](x,z)d[P - P_1 \otimes P_2](y,w) =^{3} 0$$

if and only if $P = P_1 \otimes P_2$.

The first case is usually called Hilbert Schmidt Independence Criterion (HSIC) and the second case is called Distance Covariance (Dcov).

[3] under first moment assumptions

Introduction: Distance covariance and HSIC [4]

Let $\mathfrak{I}_1 : X_1 \times X_1 \to \mathbb{R}$ and $\mathfrak{I}_2 : X_2 \times X_2 \to \mathbb{R}$ be continuous symmetric kernels (+ reasonable integrability assumptions). Then

1. For any probability P in $\mathfrak{M}(X_1 \times X_2)$ such that $P \neq \otimes P_i$

$$\int_{X_1 \times X_2} \int_{X_1 \times X_2} \mathfrak{I}_1(x_1, y_1)\mathfrak{I}_2(x_2, y_2) d[P - \otimes P_i](x) d[P - \otimes P_i](y) > 0.$$

2. For any distinct probabilities P, Q in $\mathfrak{M}(X_1 \times X_2)$ with the same marginals

$$\int_{X_1 \times X_2} \int_{X_1 \times X_2} \mathfrak{I}_1(x_1, y_1)\mathfrak{I}_2(x_2, y_2) d[P - Q](x) d[P - Q](y) > 0.$$

3. The kernels $\mathfrak{I}_1$ and $\mathfrak{I}_2$ are CND-Characteristic (up to sign change).

[4] J.C. Guella, Generalization of the HSIC and distance covariance using positive definite independent kernels

Limitations of Distance covariance and HSIC in several dimensions ($n \geq 3$)

Let $\mathfrak{I}_i : X_i \times X_i \to \mathbb{R}$, $1 \leq i \leq n$, be continuous symmetric kernels . Then (+ reasonable integrability assumptions)

1. For any probability P in $\mathfrak{M}(\mathbb{X}_n)$ such that $P \neq \otimes P_i$

$$\int_{\mathbb{X}_n} \int_{\mathbb{X}_n} \prod_{i=1}^{n} \mathfrak{I}_i(x_i, y_i) d[P - \otimes P_i](x) d[P - \otimes P_i](y) > 0$$

2. For any distinct probabilities P, Q in $\mathfrak{M}(\mathbb{X}_n)$ with the same marginals

$$\int_{\mathbb{X}_n} \int_{\mathbb{X}_n} \prod_{i=1}^{n} \mathfrak{I}_i(x_i, y_i) d[P - Q](x) d[P - Q](y) > 0$$

3. All the kernels $\mathfrak{I}_i$ are PD and ISPD (up to sign change).

The highly important trick which is heavily used in the radial case is that $\otimes_{i=1}^{n} \mu_i = M[P - \otimes_{i=1}^{n} P_i]$ whenever for at least two terms $\mu_i(X_i) = 0$.

(In progress) Positive definite independent of order 2 (PDI_2) kernels

Let $\mathfrak{M}_1(\mathbb{X}_n) := \{\mu \in \mathfrak{M}(\mathbb{X}_n), \mu(\otimes_{i=1}^{n} A_i) = 0 \text{ whenever } |\{i \mid A_i = X_i\}| = n-1\}$

Definition

A kernel $\mathfrak{I} : \mathbb{X}_n \times \mathbb{X}_n \to \mathbb{R}$ is called Positive definite independent of order 2 (PDI_2) if it is n-symmetric and for any discrete measure $\mu \in \mathfrak{M}_1(\mathbb{X}_n)$

$$\int_{\mathbb{X}_n} \int_{\mathbb{X}_n} \mathfrak{I}(x, y) d\mu(x) d\mu(y) \geq 0.$$

(In progress) Positive definite independent of order 2 (PDI$_2$) kernels

Let $\mathfrak{M}_1(\mathbb{X}_n) := \{\mu \in \mathfrak{M}(\mathbb{X}_n), \mu(\otimes_{i=1}^n A_i) = 0 \text{ whenever } |\{i \mid A_i = X_i\}| = n-1\}$

Definition

A kernel $\mathfrak{I} : \mathbb{X}_n \times \mathbb{X}_n \to \mathbb{R}$ is called Positive definite independent of order 2 (PDI$_2$) if it is n-symmetric and for any discrete measure $\mu \in \mathfrak{M}_1(\mathbb{X}_n)$

$$\int_{\mathbb{X}_n}\int_{\mathbb{X}_n} \mathfrak{I}(x,y)d\mu(x)d\mu(y) \geq 0.$$

If $n = 2$, the geometry of those kernels are a "generalization" of tensor product of Hilbert spaces ($x_{\vec{0}} = (x_1^0, x_2^0) \in X_1 \times X_2$ is fixed).

$$K^{\mathfrak{I}}(x_{\vec{1}}, x_{\vec{2}}) = \int_{X_1\times X_2}\int_{X_1\times X_2} \mathfrak{I}(u,v)d[(\delta_{x_1^1}-\delta_{x_1^0})\otimes(\delta_{x_2^1}-\delta_{x_1^0})](u)d[(\delta_{x_1^2}-\delta_{x_1^0})\otimes(\delta_{x_2^2}-\delta_{x_2^0})](v)$$

$$4\mathfrak{I}((x_1^1,x_2^1),(x_1^2,x_2^2)) - 8 \text{ terms} = \|K^{\mathfrak{I}}_{(x_1^1,x_2^1)} + K^{\mathfrak{I}}_{(x_1^2,x_2^2)} - K^{\mathfrak{I}}_{(x_1^1,x_2^2)} - K^{\mathfrak{I}}_{x_1^2,x_2^1)}\|^2_{\mathcal{H}_{K^{\mathfrak{I}}}}$$

(In progress) Positive definite independent of order 2 (PDI_2) kernels

Let $\mathfrak{M}_1(\mathbb{X}_n) := \{\mu \in \mathfrak{M}(\mathbb{X}_n), \mu(\otimes_{i=1}^n A_i) = 0 \text{ whenever } |\{i \mid A_i = X_i\}| = n-1\}$

Definition

A kernel $\mathfrak{I} : \mathbb{X}_n \times \mathbb{X}_n \to \mathbb{R}$ is called Positive definite independent of order 2 (PDI_2) if it is n-symmetric and for any discrete measure $\mu \in \mathfrak{M}_1(\mathbb{X}_n)$

$$\int_{\mathbb{X}_n}\int_{\mathbb{X}_n} \mathfrak{I}(x,y)d\mu(x)d\mu(y) \geq 0.$$

If $n = 2$, the geometry of those kernels are a "generalization" of tensor product of Hilbert spaces ($x_{\vec{0}} = (x_1^0, x_2^0) \in X_1 \times X_2$ is fixed).

$$K^{\mathfrak{I}}(x_{\vec{1}}, x_{\vec{2}}) = \int_{X_1\times X_2}\int_{X_1\times X_2} \mathfrak{I}(u,v)d[(\delta_{x_1^1}-\delta_{x_1^0})\otimes(\delta_{x_2^1}-\delta_{x_1^0})](u)d[(\delta_{x_1^2}-\delta_{x_1^0})\otimes(\delta_{x_2^2}-\delta_{x_2^0})](v)$$

$$4\mathfrak{I}((x_1^1, x_2^1),(x_1^2, x_2^2)) - 8 \text{ terms} = \|K^{\mathfrak{I}}_{(x_1^1,x_2^1)} + K^{\mathfrak{I}}_{(x_1^2,x_2^2)} - K^{\mathfrak{I}}_{(x_1^1,x_2^2)} - K^{\mathfrak{I}}_{x_1^2,x_2^1)}\|^2_{\mathcal{H}_{K^{\mathfrak{I}}}}$$

If $n \geq 3$, (especially if $n \geq 5$), they are much more difficult to understand. $K^{\mathfrak{I}}$ is related to $\delta_{x_{\vec{1}}} - \sum_{i=1}^n \delta_{x_{e_i}} + (n-1)\delta_{x_{\vec{0}}}$, the reverse equation has $\sum_{k=2}^n \binom{n}{k} 2^{n-k}$ at the right side.

However, radial PDI kernels in all dimensions (that is, generalizations of the famous Schoenberg's results) are more well behaved.

(In progress) Limitations of Distance covariance in several dimensions ($n \geq 3$)

Let $F^1, \ldots, F^\ell$ be a disjoint family of subsets of $\{1, \ldots, n\}$ whose union is the entire set, where $\ell \geq 2$ and $|F^k|$-symmetric kernels $\mathfrak{I}_k : \mathbb{X}_{F^k} \times \mathbb{X}_{F^k} \to \mathbb{R}$, $1 \leq k \leq \ell$ (+ reasonable technical conditions). Then the kernel

$$\mathfrak{I}(x_{\vec{1}}, x_{\vec{2}}) := \prod_{k=1}^{\ell} \mathfrak{I}_k(x_{\vec{1}(F^k)}, x_{\vec{2}(F^k)})$$

is PDI_2 if and only if one of the following conditions is satisfied (up to sign change)

(i) $\ell = 2$ and $|F^1| = 1$: $\mathfrak{I}_1$ is PD and $-\mathfrak{I}_2$ is CND.

(ii) All kernels $\mathfrak{I}_i$ are PD.

Theorem

Let $g : [0,\infty)^n \to \mathbb{R}$ be a continuous function. The following conditions are equivalent:

(*i*) *The kernel $g(\|x_1 - y_1\|^2, \dots, \|x_n - y_n\|^2), \quad x_i, y_i \in \mathbb{R}^d$ is PD for every $d \in \mathbb{N}$.*

(*ii*) *The function can be represented as*

$$g(t) = \int_{[0,\infty)^n} e^{-r\cdot t} d\eta(r_1, \dots, r_n)$$

where the measure $\eta \in \mathfrak{M}([0,\infty)^n)$ is nonnegative. Further, the representation is unique.

(*iii*) *The function g is completely monotone in $(0,\infty)^n$, that is $g \in C^\infty((0,\infty)^n)$ and $(-1)^{|\alpha|}[\partial^\alpha g](t) \geq 0$ for any $\alpha \in (\mathbb{Z}_+)^n$ and $t \in (0,\infty)^n$.*

ISPD if and only if $\eta((0,\infty)^n) > 0$.

Equivalence between ii and iii:"Harmonic Analysis and the Theory of Probability, Bochner 1955".

(In progress) Multivariate Radial CND kernels

Theorem

Let $g : [0, \infty)^n \to \mathbb{R}$ be a continuous function such that $g(0) = 0$. The following conditions are equivalent:

(*i*) *The kernel $g(\|x_1 - y_1\|^2, \ldots, \|x_n - y_n\|^2)$, $\quad x_i, y_i \in \mathbb{R}^d$ is CND for every $d \in \mathbb{N}$.*

(*ii*) *The function can be represented as*

$$g(t) = \sum_{i=1}^{n} a_i t_i + \int_{[0,\infty)^n \setminus \{0\}} (1 - e^{-r \cdot t}) \frac{1 + \sum_{i=1}^{n} r_i}{\sum_{i=1}^{n} r_i} d\eta(r_1, \ldots, r_n)$$

where the measure $\eta \in \mathfrak{M}([0, \infty)^n)$ and the scalars a_i are nonnegative. Further, the representation is unique.

(*iii*) *The function g is a Bernstein function of order 1 in $(0, \infty)^n$, that is $g \in C^{\infty}((0, \infty)^n)$ and $\partial^{e_i} g$ is completely monotone for every $1 \leq i \leq n$.*

Equivalence between ii and iii: "Properties of Bernstein Functions of Several Complex Variables A. R. Mirotin, 2013".

CND-Characteristic if and only if $\eta((0, \infty)^n) > 0$.

We use the elementary symmetric polynomials

$$p_2^n(r_1, \ldots, r_n) := \sum_{1 \leq i < j \leq n} r_i r_j, \quad p_1^n(r_1, \ldots, r_n) = \sum_{1 \leq i \leq n} r_i.$$

The zeroes of p_2^n in the set $[0, \infty)^n$ is the set

$$\partial_1^n := \bigcup_{F \subset \{1, \ldots, n\}, |F|=1} \{\lambda_i e_i, \quad \lambda_i \geq 0, 1 \leq i \leq n\}$$

(In progress) Multivariate Radial PDI$_2$ kernels

Theorem

Let $n \geq 2$ and $g : [0, \infty)^n \to \mathbb{R}$ be a continuous function that is zero in ∂_1^n. The following conditions are equivalent:

(*i*) *For any $d \in \mathbb{N}$ and discrete probability P in $(\mathbb{R}^d)^n$, with marginals P_i in $\mathbb{R}^d$, it holds that*

$$\int_{(\mathbb{R}^d)^n} \int_{(\mathbb{R}^d)^n} g(\|x_1 - y_1\|^2, \ldots, \|x_n - y_n\|^2) d[P - \otimes_{i=1}^n P_i](x) d[P - \otimes_{i=1}^n P_i](y) \geq 0.$$

(*ii*) *The kernel $g(\|x_1 - y_1\|^2, \ldots, \|x_n - y_n\|^2), \quad x_i, y_i \in \mathbb{R}^d$ is PDI$_2$ for every $d \in \mathbb{N}$.*

(*iii*) *The function can be represented as*

$$g(t) = \sum_{i \neq j} t_i \psi^{i,j}(t_j) + \sum_{i<j} a^{i,j} t_i t_j + \int_{[0,\infty)^n \setminus \partial_1^n} \left(e^{-r \cdot t} - \sum_{i=1}^{n} e^{-r_i t_i} + n - 1 \right) \frac{1 + p_1^n(r) + p_2^n(r)}{p_2^n(r)} d\eta(r)$$

where the measure $\eta \in \mathfrak{M}([0, \infty)^n \setminus \partial_1^n)$ and the scalars $a^{i,j}$ are nonnegative and the functions $\psi^{i,j}$ are Bernstein. Further, the representation is unique.

(*iv*) *The function $g(t)$ is a Bernstein function of order 2 in $(0, \infty)^n$, that is $g \in C^\infty((0, \infty)^n)$ and $\partial^{e_i + e_j} g$ is completely monotone for every $1 \leq i < j \leq n$.*

PDI$_2$-Characteristic if and only if $\eta((0, \infty)^n) > 0$.

Distance Multivariance

Related to recent works of Bjorn Bottcher, Martin Keller-Ressel and Rene L. Schilling.

Let $\mathfrak{I}_i : X_i \times X_i \to \mathbb{R}$, $1 \le i \le n$, be continuous symmetric kernels. Then (+ reasonable integrability assumptions)

1. For any probability P in $\mathfrak{M}(\mathbb{X}_n)$ such that $\Delta_S^n P \neq 0$ (Streitberg interaction)

$$\int_{\mathbb{X}_n} \int_{\mathbb{X}_n} \prod_{i=1}^{n} \mathfrak{I}_i(x_i, y_i) d[\Delta_S^n P](x) d[\Delta_S^n P](y) > 0.$$

   Similarly for the Lancaster interaction.

2. For any distinct probabilities P, Q in $\mathfrak{M}(\mathbb{X}_n)$ with the same "complemented" marginals ($P(\otimes_{i=1}^n A_i) = Q(\otimes_{i=1}^n A_i)$ whenever $|\{i \mid A_i = X_i\}| \ge 1$)

$$\int_{\mathbb{X}_n} \int_{\mathbb{X}_n} \prod_{i=1}^{n} \mathfrak{I}_i(x_i, y_i) d[P - Q](x) d[P - Q](y) > 0.$$

3. All the kernels $\mathfrak{I}_i$ are CND-Characteristic (up to sign change).

Distance Multivariance

Related to recent works of Bjorn Bottcher, Martin Keller-Ressel and Rene L. Schilling.

Let $\mathfrak{I}_i : X_i \times X_i \to \mathbb{R}$, $1 \leq i \leq n$, be continuous symmetric kernels. Then (+ reasonable integrability assumptions)

1. For any probability P in $\mathfrak{M}(\mathbb{X}_n)$ such that $\Delta_S^n P \neq 0$ (Streitberg interaction)

$$\int_{\mathbb{X}_n} \int_{\mathbb{X}_n} \prod_{i=1}^n \mathfrak{I}_i(x_i, y_i) d[\Delta_S^n P](x) d[\Delta_S^n P](y) > 0.$$

   Similarly for the Lancaster interaction.

2. For any distinct probabilities P, Q in $\mathfrak{M}(\mathbb{X}_n)$ with the same "complemented" marginals ($P(\otimes_{i=1}^n A_i) = Q(\otimes_{i=1}^n A_i)$ whenever $|\{i \mid A_i = X_i\}| \geq 1$)

$$\int_{\mathbb{X}_n} \int_{\mathbb{X}_n} \prod_{i=1}^n \mathfrak{I}_i(x_i, y_i) d[P - Q](x) d[P - Q](y) > 0.$$

3. All the kernels $\mathfrak{I}_i$ are CND-Characteristic (up to sign change).

This result leads to a new type o kernel (Positive definite independent of order n, PDI_n), which has an easier mathemathical structure (generalization of an n-tensor product).

Distance Multivariance

Related to recent works of Bjorn Bottcher, Martin Keller-Ressel and Rene L. Schilling.

Let $\mathfrak{I}_i : X_i \times X_i \to \mathbb{R}$, $1 \leq i \leq n$, be continuous symmetric kernels. Then (+ reasonable integrability assumptions)

1. For any probability P in $\mathfrak{M}(\mathbb{X}_n)$ such that $\Delta_S^n P \neq 0$ (Streitberg interaction)

$$\int_{\mathbb{X}_n}\int_{\mathbb{X}_n}\prod_{i=1}^{n}\mathfrak{I}_i(x_i, y_i)d[\Delta_S^n P](x)d[\Delta_S^n P](y) > 0.$$

   Similarly for the Lancaster interaction.

2. For any distinct probabilities P, Q in $\mathfrak{M}(\mathbb{X}_n)$ with the same "complemented" marginals ($P(\otimes_{i=1}^n A_i) = Q(\otimes_{i=1}^n A_i)$ whenever $|\{i \mid A_i = X_i\}| \geq 1$)

$$\int_{\mathbb{X}_n}\int_{\mathbb{X}_n}\prod_{i=1}^{n}\mathfrak{I}_i(x_i, y_i)d[P - Q](x)d[P - Q](y) > 0.$$

3. All the kernels $\mathfrak{I}_i$ are CND-Characteristic (up to sign change).

This result leads to a new type o kernel (Positive definite independent of order n, PDI_n), which has an easier mathemathical structure (generalization of an n-tensor product). Intermediate cases between PDI_2 and PDI_n are possible to analyse, but its difficult to handle the combinatorial burden of its terminology.

Let
$\mathfrak{M}_{n-1}(\mathbb{X}_n) := \{\mu \in \mathfrak{M}(\mathbb{X}_n), \mu(\otimes_{i=1}^n A_i) = 0 \text{ whenever } |\{i \mid A_i = X_i\}| = 1\} \subset \mathfrak{M}_1(\mathbb{X}_n)$

Definition

A kernel $\mathfrak{I} : \mathbb{X}_n \times \mathbb{X}_n \to \mathbb{R}$ is called Positive definite independent of order n (PDI$_n$) if it is n-symmetric and for any discrete measure $\mu \in \mathfrak{M}_n(\mathbb{X}_n)$

$$(-1)^n \int_{\mathbb{X}_n} \int_{\mathbb{X}_n} \mathfrak{I}(x, y) d\mu(x) d\mu(y) \geq 0.$$

$$K^{\mathfrak{I}}(x_{\vec{1}}, x_{\vec{2}}) = \int_{\mathbb{X}_n} \int_{\mathbb{X}_n} (-1)^n \mathfrak{I}(u, v) d[\otimes_{i=1}^n (\delta_{x_i^1} - \delta_{x_i^0})](u) d[\otimes_{i=1}^n (\delta_{x_i^2} - \delta_{x_i^0})](v)$$

$$2^n \mathfrak{I}(x_{\vec{1}}, x_{\vec{2}}) - \text{ huge amount of terms} = \| \sum_{\alpha \in \mathbb{N}_2^n} (-1)^{|\alpha|} K^{\mathfrak{I}}_{x_\alpha} \|^2_{\mathcal{H}_{K^{\mathfrak{I}}}}$$

(In progress) Multivariate Radial PDI$_n$ kernels

Theorem

Let $g : [0,\infty)^n \to \mathbb{R}$ be a continuous function such that g is zero at the border of $[0,\infty)^n$. The following conditions are equivalent:

(i) *The kernel $g(\|x_1 - y_1\|^2, \ldots, \|x_n - y_n\|^2), \quad x_i, y_i \in \mathbb{R}^d$ is PDI$_n$ for every $d \in \mathbb{N}$.*

(ii) *The function can be represented as*

$$g(t) = \int_{[0,\infty)^n} \left[\prod_{i=1}^{n}(1 - e^{-r_i t_i})\right] \frac{\prod_{i=1}^{n}(1+r_i)}{\prod_{i=1}^{n} r_i} d\eta(r_1, \ldots, r_n)$$

where the measure $\eta \in \mathfrak{M}([0,\infty)^n)$ is nonnegative. Further, the representation is unique.

(iii) *The function g is a Bernstein function of order n in $(0,\infty)^n$, that is $g \in C^\infty((0,\infty)^n)$ and $\partial^{\vec{1}} g$ is completely monotone.*

PDI$_n$-Characteristic if and only if $\eta((0,\infty)^n) > 0$.

Thank you!