HSIC, A Measure of Statistical Independence?*

Zoltán Szabó[1], Bharath K. Sriperumbudur[2]

[1]CMAP, École Polytechnique [2]Department of Statistics, Pennsylvania State University

Quick Summary

- Focus:
  - measuring independence,
  - on kernel-endowed domains.
- Dependency measure:
  - Hilbert-Schmidt independence criterion (HSIC).
- Goal: To understand
  - when HSIC is an independence measure,
  - more generally when a tensor product kernel is 'characteristic'.

Independence, Kernels

- $X = (X_m)_{m=1}^M \in \mathcal{X} = \times_{m=1}^M \mathcal{X}_m$: random variable.
- Task: measure the statistical dependence of X_m-s.
- Alternatively: If $X \sim \mathbb{P}$, $X_m \sim \mathbb{P}_m$,

$$\mathbb{P} \stackrel{?}{=} \otimes_{m=1}^M \mathbb{P}_m.$$

- Assumption: $\mathcal{X}_m$-s are kernel-enriched
  - Examples: trees, graphs, strings, time series, hidden Markov models, sets, fuzzy domains, distributions, groups $\xrightarrow{\text{spec.}}$ permutations.

Distribution Representation

Mean embedding, maximum mean discrepancy, HSIC:

$$\mu_k(\mathbb{P}) = \int_{\mathcal{X}} \underbrace{\varphi(x)}_{k(\cdot,x)} \, d\mathbb{P}(x) \in \mathcal{H}_k,$$

$$\text{MMD}_k(\mathbb{P}, \mathbb{Q}) = \|\mu_k(\mathbb{P}) - \mu_k(\mathbb{Q})\|_{\mathcal{H}_k},$$

$$\text{HSIC}_k(\mathbb{P}) = \text{MMD}_k\left(\mathbb{P}, \otimes_{m=1}^M \mathbb{P}_m\right), k(x, x') = \prod_{m=1}^M k_m(x_m, x'_m).$$

*Data Learning and Inference (DALI), Lanzarote, Canary Islands, Spain, 3-5 April, 2018.

Central: Characteristic Property

- Mean embedding, MMD, HSIC: numerous applications. Review [4].
- MMD: k is called
  - characteristic: if $\text{MMD}_k(\mathbb{P}, \mathbb{Q}) = 0 \Leftrightarrow \mathbb{P} = \mathbb{Q}$.
  - universal: injectivity on finite signed measures.
- HSIC: $k = \otimes_{m=1}^M k_m$ is $\mathcal{I}$-characteristic if

$$\text{HSIC}_k(\mathbb{P}) = 0 \Leftrightarrow \mathbb{P} = \otimes_{m=1}^M \mathbb{P}_m.$$

$\otimes_{m=1}^M k_m$: universal $\Rightarrow$ characteristic $\Rightarrow$ $\mathcal{I}$-characteristic.

- **Wanted**: Converse? Description in terms of k_m-s!

Existing Results: $M = 2$

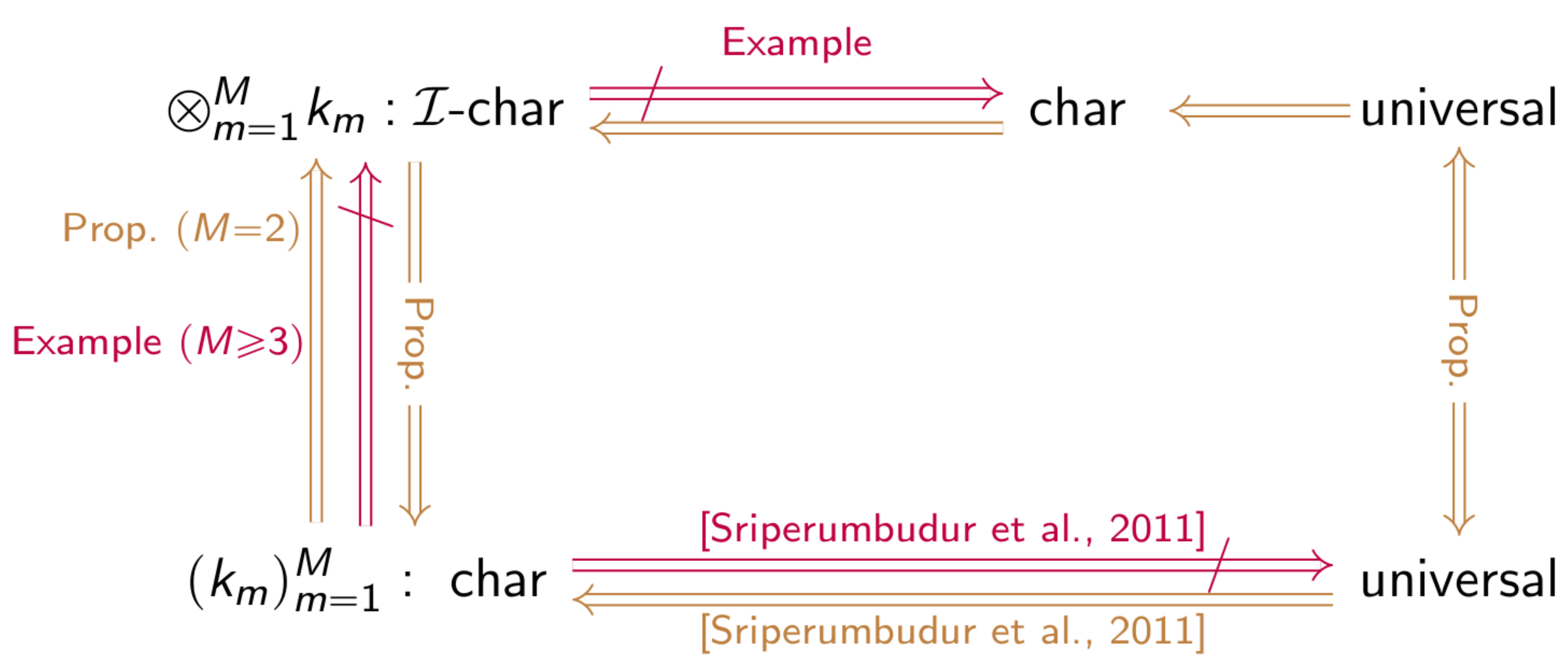
[1, 2]; distance covariance [3, 5]:

$k_1 \& k_2$: universal $\Rightarrow k_1 \otimes k_2$: universal ($\Rightarrow \mathcal{I}$-char).

$k_1 \& k_2$: characteristic $\Leftrightarrow k_1 \otimes k_2$: $\mathcal{I}$-characteristic.

Our Results [7]

Visual summary:



Key: $\mathcal{F}$-positive definiteness (e.g., $\mathcal{F} = \{\mathbb{P} - \otimes_{m=1}^M \mathbb{P}_m\}$)

$$\|\mu_k(\mathbb{F})\|_{\mathcal{H}_k}^2 > 0, \quad \forall \mathbb{F} \in \mathcal{F} \backslash \{0\}.$$

- Characteristic property:
  - $\otimes_{m=1}^M k_m$: characteristic $\Rightarrow (k_m)_{m=1}^M$ are characteristic.
  - $\not\Leftarrow [|\mathcal{X}_m| = 2, k_m(x, x') = 2\delta_{x,x'} - 1]$.
- $\mathcal{I}$-characteristic property:
  - k_1, k_2: characteristic $\Rightarrow k_1 \otimes k_2$: $\mathcal{I}$-characteristic.
  - $\Leftarrow$ for $\forall M \geq 2$.
  - k_1, k_2, k_3: characteristic $\not\Rightarrow \otimes_{m=1}^3 k_m$: $\mathcal{I}$-characteristic [Ex.].
- $\mathcal{X}_m = \mathbb{R}^{d_m}$, k_m: continuous, shift-invariant, bounded:
  - $(k_m)_{m=1}^M$: characteristic $\Leftrightarrow \otimes_{m=1}^M k_m$: $\mathcal{I}$-characteristic $\Leftrightarrow$ $\otimes_{m=1}^M k_m$: characteristic.
- Universality:
  - $\otimes_{m=1}^M k_m$: universal $\Leftrightarrow (k_m)_{m=1}^M$: universal.

Acknowledgements. A part of the work was carried out while BKS was visiting ZSz at CMAP, École Polytechnique. BKS is supported by NSF-DMS-1713011 and also thanks CMAP and DSI for their generous support. ZSz is supported by DSI.

References

[1] G. Blanchard, G. Lee, and C. Scott. Generalizing from several related classification tasks to a new unlabeled sample. In *NIPS*, pages 2178–2186, 2011.

[2] A. Gretton. A simpler condition for consistency of a kernel independence test. Technical report, UCL, 2015. (http://arxiv.org/abs/1501.06103).

[3] R. Lyons. Distance covariance in metric spaces. *The Annals of Probability*, 41:3284–3305, 2013.

[4] K. Muandet, K. Fukumizu, B. K. Sriperumbudur, and B. Schölkopf. Kernel mean embedding of distributions: A review and beyond. *Foundations and Trends in Machine Learning*, 10(1-2):1–141, 2017.

[5] D. Sejdinovic, B. K. Sriperumbudur, A. Gretton, and K. Fukumizu. Equivalence of distance-based and RKHS-based statistics in hypothesis testing. *Annals of Statistics*, 41:2263–2291, 2013.

[6] Bharath K. Sriperumbudur, Kenji Fukumizu, and Gert R. G. Lanckriet. Universality, characteristic kernels and RKHS embedding of measures. *JMLR*, 12:2389–2410, 2011.

[7] Z. Szabó and B. K. Sriperumbudur. Characteristic and universal tensor product kernels. Technical report, 2017. (http://arxiv.org/abs/1708.08157).