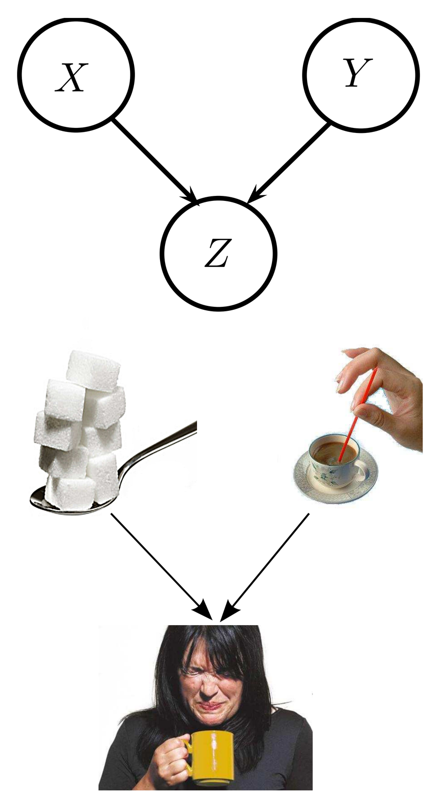
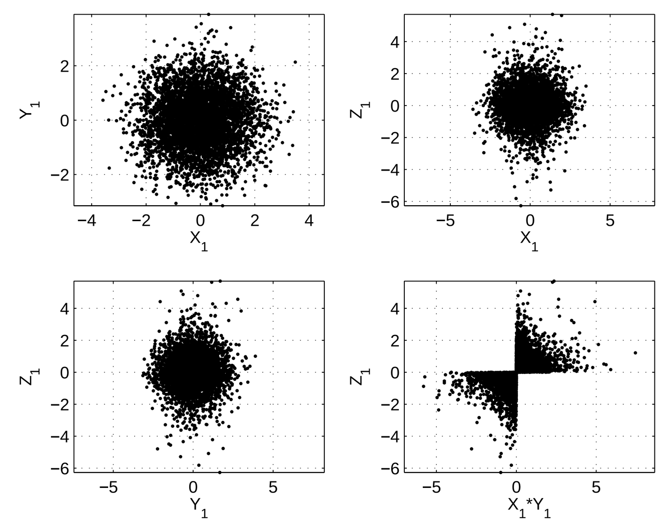
A Kernel Test for Three Variable Interactions

Dino Sejdinovic[1] **Arthur Gretton**[1] **Wicher Bergsma**[2]

[1]Gatsby Computational Neuroscience Unit, CSML, University College London, [2]Department of Statistics, London School of Economics

Motivation: How to detect V-structures with pairwise weak (or nonexistent) association?





Example A: $X \perp\!\!\!\perp Y, Y \perp\!\!\!\perp Z, X \perp\!\!\!\perp Z$

- $X_1, Y_1 \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$
- $Z_1 | X_1, Y_1 \sim \text{sign}(X_1 Y_1) Exp(\frac{1}{\sqrt{2}})$
- $X_{2:p}, Y_{2:p}, Z_{2:p} \overset{i.i.d.}{\sim} \mathcal{N}(0, \mathbf{I}_{p-1})$

Embeddings of probability measures into RKHS

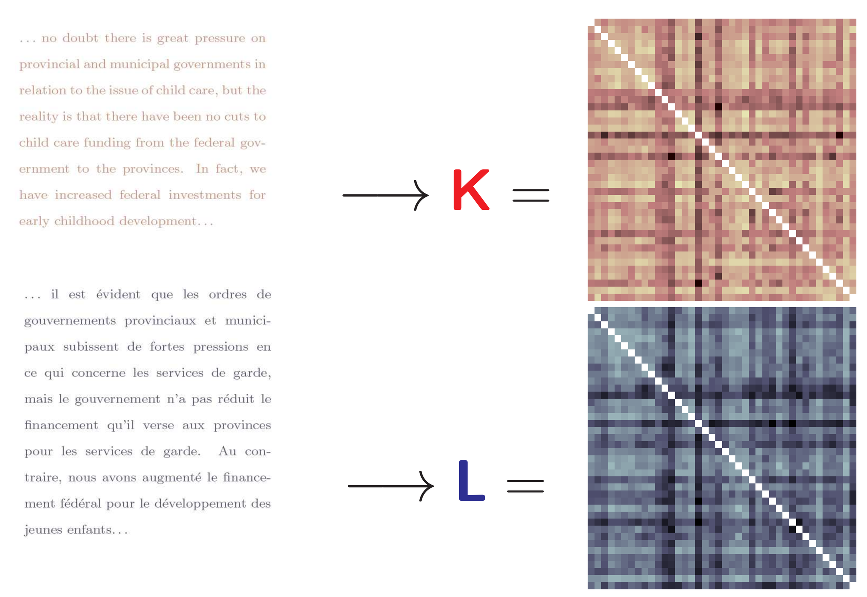
- **Feature map**: $z \mapsto k(\cdot, z) \in \mathcal{H}_k$ instead of $z \mapsto (\varphi_1(z), \ldots, \varphi_s(z)) \in \mathbb{R}^s$
- $\langle k(\cdot,z), k(\cdot,w)\rangle_{\mathcal{H}_k} = k(z,w)$ ← inner products easily **computed**
- **Kernel embedding**: $P \mapsto \mu_k(P) = \mathbb{E}_{Z\sim P} k(\cdot, Z) \in \mathcal{H}_k$ instead of $P \mapsto (\mathbb{E}\varphi_1(Z), \ldots, \mathbb{E}\varphi_s(Z)) \in \mathbb{R}^s$
- $\langle \mu_k(P), \mu_k(Q)\rangle_{\mathcal{H}_k} = \mathbb{E}_{Z\sim P, W\sim Q} k(Z,W)$ ← inner products easily **estimated**

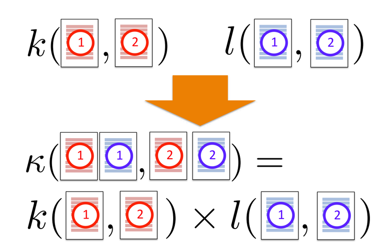
Pairwise RKHS-based independence test

- **Hilbert-Schmidt Independence Criterion** (Gretton et al 2005, 2008; Smola et al 2007):

$$\left\| \mu_\kappa(\hat{P}_{XY}) - \mu_\kappa(\hat{P}_X \hat{P}_Y) \right\|^2_{\mathcal{H}_\kappa} = \frac{1}{n^2} (H\mathbf{K}H \circ H\mathbf{L}H)_{++}$$

- Powerful nonparametric independence tests that generalize (Sejdinovic et al 2013) distance covariance (**dCov**) of Szekely et al (2007)





- **HSIC** $= \frac{1}{n^2} \langle H\mathbf{K}H, H\mathbf{L}H \rangle$
- $H = I - \frac{1}{n}\mathbf{1}\mathbf{1}^\top$ (centering matrix)
- $A_{++} = \sum_{i=1}^n \sum_{j=1}^n A_{ij}$

Hilbert-Schmidt norm of the covariance operator in the feature space

How to detect a V-structure using the existing RKHS-based tests?

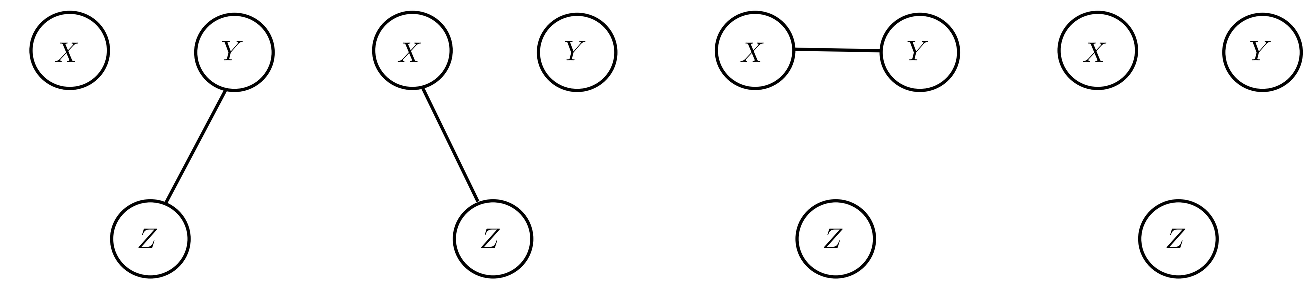
Assuming $X \perp\!\!\!\perp Y$ has been established:

- **Conditional independence test**: $\mathbf{H_0}: X \perp\!\!\!\perp Y | Z$ (Zhang et al 2011) or
- **Factorisation test**: $\mathbf{H_0}: (X,Y) \perp\!\!\!\perp Z \vee (X,Z) \perp\!\!\!\perp Y \vee (Y,Z) \perp\!\!\!\perp X$ (multiple standard two-variable tests)
  - compute p-values for each of the marginal tests
  - apply Holm-Bonferroni sequentially rejective correction (Holm 1979)

Capturing factorisations directly: Interaction Measures (Lancaster, 1969)

Interaction measure of $(X_1, \ldots, X_D) \sim P_X$ is a signed measure ΔP that vanishes whenever P can be factorised in a non-trivial way as a product of its (possibly multivariate) marginals.

- $D = 2$: $\Delta_L P = P_{XY} - P_X P_Y$
- $D = 3$:



$\Delta_L P = P_{XYZ} - P_X P_{YZ} - P_{XZ} P_Y - P_{XY} P_Z + 2 P_X P_Y P_Z$

Inner product estimators

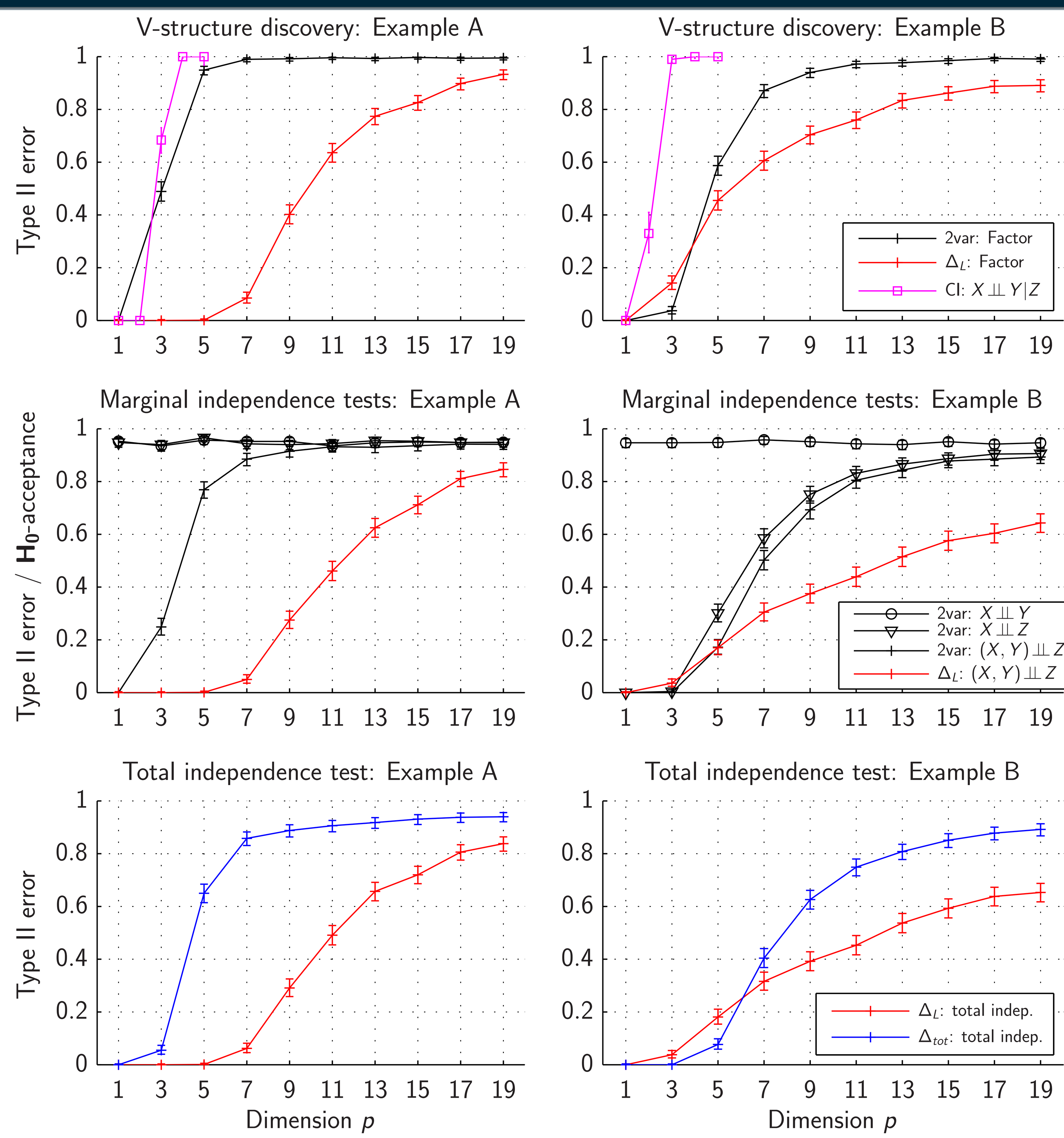
$\nu \backslash \nu'$	P_{XYZ}	$P_{XY}P_Z$	$P_{XZ}P_Y$	$P_{YZ}P_X$	$P_X P_Y P_Z$
P_{XYZ}	$(\mathbf{K}\circ\mathbf{L}\circ\mathbf{M})_{++}$	$((\mathbf{K}\circ\mathbf{L})\mathbf{M})_{++}$	$((\mathbf{K}\circ\mathbf{M})\mathbf{L})_{++}$	$((\mathbf{M}\circ\mathbf{L})\mathbf{K})_{++}$	$tr(\mathbf{K}_+\circ\mathbf{L}_+\circ\mathbf{M}_+)$
$P_{XY}P_Z$		$(\mathbf{K}\circ\mathbf{L})_{++}\mathbf{M}_{++}$	$(\mathbf{MKL})_{++}$	$(\mathbf{KLM})_{++}$	$(\mathbf{KL})_{++}\mathbf{M}_{++}$
$P_{XZ}P_Y$			$(\mathbf{K}\circ\mathbf{M})_{++}\mathbf{L}_{++}$	$(\mathbf{KML})_{++}$	$(\mathbf{KM})_{++}\mathbf{L}_{++}$
$P_{YZ}P_X$				$(\mathbf{L}\circ\mathbf{M})_{++}\mathbf{K}_{++}$	$(\mathbf{LM})_{++}\mathbf{K}_{++}$
$P_X P_Y P_Z$					$\mathbf{K}_{++}\mathbf{L}_{++}\mathbf{M}_{++}$

Table: V-statistic estimators of $\langle \mu_\kappa \nu, \mu_\kappa \nu' \rangle_{\mathcal{H}_\kappa}$, with $\kappa = k \otimes l \otimes m$

$$\left\| \mu_\kappa\left(\Delta_L \hat{P}\right) \right\|^2_{\mathcal{H}_\kappa} = \frac{1}{n^2} (H\mathbf{K}H \circ H\mathbf{L}H \circ H\mathbf{M}H)_{++}$$

Norm of the empirical joint central moment in the feature space

Results: gaussian kernel with median distance bandwidth, $n = 500$



What kind of interactions does the Lancaster statistic capture?

If $\kappa = k \otimes l \otimes m$ is **integrally strictly positive definite** (Sriperumbudur, 2010), then

$$\|\mu_\kappa(\Delta_L P)\|_{\mathcal{H}_\kappa} = 0 \Leftrightarrow \Delta_L P = 0.$$

Theorem: $\|\mu_\kappa(\Delta_L P)\|_{\mathcal{H}_\kappa} = 0 \Leftrightarrow \mathbb{E}_{XYZ}[(f(X) - \mathbb{E}f)(g(Y) - \mathbb{E}g)(h(Z) - \mathbb{E}h)] = 0$ for all $f \in \mathcal{H}_k$, $g \in \mathcal{H}_l$, $h \in \mathcal{H}_m$.

However, $\Delta_L P = 0 \not\Rightarrow (X,Y) \perp\!\!\!\perp Z \vee (X,Z) \perp\!\!\!\perp Y \vee (Y,Z) \perp\!\!\!\perp X$

Counterexample: X, Y, Z binary and $P(x,y,z) = \begin{cases} 0.2 & x = y = z, \\ 0.1 & \text{otherwise.} \end{cases}$

Example B

- **Pairwise dependence present, but weaker**
- $X_1, Y_1 \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$
- $Z_1 | X_1, Y_1 \sim \begin{cases} X_1^2 + \mathcal{N}(0, 0.01), & w.p.\ 1/3, \\ Y_1^2 + \mathcal{N}(0, 0.01), & w.p.\ 1/3, \\ X_1 Y_1 + \mathcal{N}(0, 0.01), & w.p.\ 1/3. \end{cases}$
- $X_{2:p}, Y_{2:p}, Z_{2:p} \overset{i.i.d.}{\sim} \mathcal{N}(0, \mathbf{I}_{p-1})$

Permutation tests based on Lancaster statistic

- Marginal independence test, e.g., $\mathbf{H_0}: (X,Y) \perp\!\!\!\perp Z$: estimate null distribution using $(X^{(i)}, Y^{(i)}, Z^{(\sigma i)})_{i=1}^n$, $\sigma \in S_n$
- Factorisation test: multiple marginal independence tests with Holm-Bonferroni correction
- Total independence test $\mathbf{H_0}: P_{XYZ} = P_X P_Y P_Z$: estimate null distribution using $(X^{(i)}, Y^{(\tau i)}, Z^{(\sigma i)})_{i=1}^n$, $\sigma, \tau \in S_n$

Total independence test $\mathbf{H_0}: P_\mathbf{X} = \prod_{i=1}^D P_{X_i}$

For $(X_1, \ldots, X_D) \sim P_\mathbf{X}$, and $\kappa = \bigotimes_{i=1}^D k^{(i)}$:

$$\Bigg\| \mu_\kappa \underbrace{\left(\hat{P}_\mathbf{X} - \prod_{i=1}^D \hat{P}_{X_i} \right)}_{\Delta_{tot}\hat{P}} \Bigg\|^2_{\mathcal{H}_\kappa} = \frac{1}{n^2} \sum_{a=1}^n \sum_{b=1}^n \prod_{i=1}^D K_{ab}^{(i)} - \frac{2}{n^{D+1}} \sum_{a=1}^n \prod_{i=1}^D \sum_{b=1}^n K_{ab}^{(i)} + \frac{1}{n^{2D}} \prod_{i=1}^D \sum_{a=1}^n \sum_{b=1}^n K_{ab}^{(i)}.$$

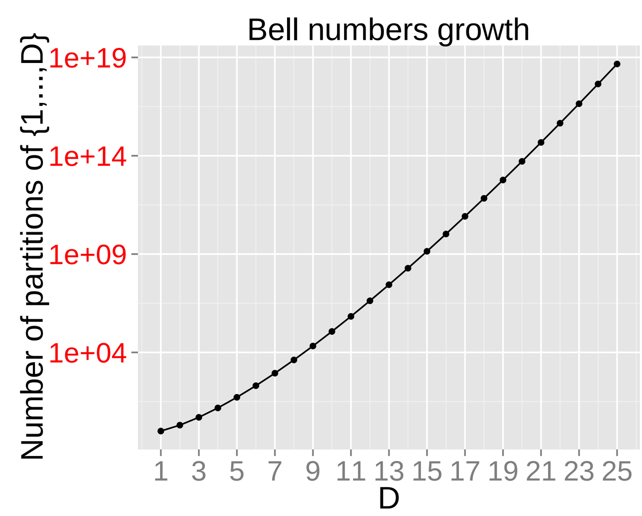
Coincides with the test proposed by Kankainen (1995) using empirical characteristic functions: similar relationship to that between **dCov** and **HSIC** (Sejdinovic et al, 2013)

Interaction for $D \geq 4$

- Interaction measure valid for all D (Streitberg, 1990):

$$\Delta_S P = \sum_\pi (-1)^{|\pi|-1} (|\pi| - 1)! J_\pi P$$

- For a partition π, J_π associates to the joint the corresponding factorization, e.g., $J_{13|2|4} P = P_{X_1 X_3} P_{X_2} P_{X_4}$.
- Overal expression does not collapse and summing over all partitions is required.



$\mu_\kappa(\Delta_L P)$: **joint central moment in the RKHS** vs. $\mu_\kappa(\Delta_S P)$: **joint cumulants in the RKHS**

References

- H.O. Lancaster *The Chi-Squared Distribution* - **Ch. XII**. Wiley, London, 1969.
- B. Streitberg **Lancaster interactions revisited**. *Ann. Stat.* **18**(4): 1878-1885, 1990.
- A. Kankainen.*Consistent Testing of Total Independence Based on the Empirical Characteristic Function*. University of Jyväskylä, 1995.
- A. Gretton, K. Fukumizu, C.-H. Teo, L. Song, B. Schölkopf, and A. Smola. **A kernel statistical test of independence.** *Advances in Neural Information Processing Systems* **20**: 585-592, 2008.
- B. Sriperumbudur, A. Gretton, K. Fukumizu, G. Lanckriet, and B. Schölkopf. **Hilbert space embeddings and metrics on probability measures.** *J. Mach. Learn. Res.* **11**: 1517-1561, 2010.
- K. Zhang, J. Peters, D. Janzing, and B. Schölkopf. **Kernel-based conditional independence test and application in causal discovery.** In *Uncertainty in Artificial Intelligence* 804-813, 2011.
- D. Sejdinovic, B. Sriperumbudur, A. Gretton, and K. Fukumizu, **Equivalence of distance-based and RKHS-based statistics in hypothesis testing.** *Ann. Stat.* **41**(5): 2263-2291, 2013.

This work was funded by the Gatsby Charitable Foundation.

Contact: dino@gatsby.ucl.ac.uk