A Kernel Test for Three Variable Interactions

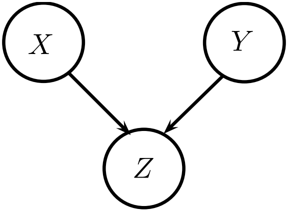
Dino Sejdinovic[1], Arthur Gretton[1], Wicher Bergsma[2]

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

NIPS, 07 Dec 2013

Detecting a higher order interaction

- How to detect V-structures with pairwise weak (or nonexistent) dependence?



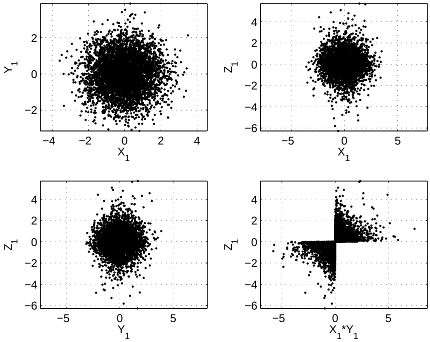
Detecting a higher order interaction

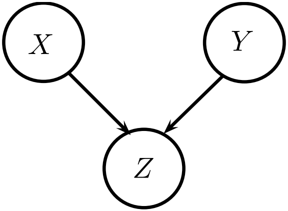
- How to detect V-structures with pairwise weak (or nonexistent) dependence?

Detecting a higher order interaction

- How to detect V-structures with pairwise weak (or nonexistent) dependence?

- $X \perp\!\!\!\perp Y$, $Y \perp\!\!\!\perp Z$, $X \perp\!\!\!\perp Z$





- $X, Y \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$,
- $Z | X, Y \sim \mathrm{sign}(XY) Exp(\frac{1}{\sqrt{2}})$

Detecting pairwise dependence

- How to detect dependence in a **non-Euclidean / structured** domain?

X_1: Honourable senators, I have a question for the Leader of the Government in the Senate with regard to the support funding to farmers that has been announced. Most farmers have not received any money yet.

X_2: No doubt there is great pressure on provincial and municipal governments in relation to the issue of child care, but the reality is that there have been no cuts to child care funding from the federal government to the provinces. In fact, we have increased federal investments for early childhood development.

. . .

$$\overset{?}{\Longleftrightarrow}$$

Y_1: Honorables sénateurs, ma question s'adresse au leader du gouvernement au Sénat et concerne l'aide financiére qu'on a annoncée pour les agriculteurs. La plupart des agriculteurs n'ont encore rien reçu de cet argent.

Y_2: Il est évident que les ordres de gouvernements provinciaux et municipaux subissent de fortes pressions en ce qui concerne les services de garde, mais le gouvernement n'a pas réduit le financement qu'il verse aux provinces pour les services de garde. Au contraire, nous avons augmenté le financement fédéral pour le développement des jeunes enfants.

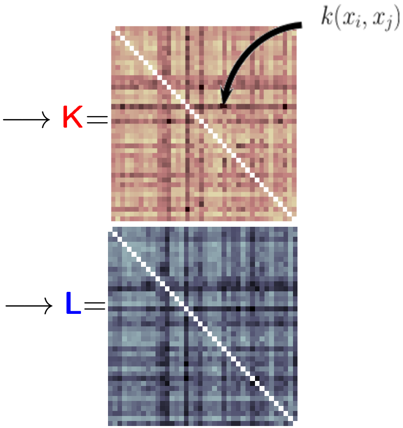
. . .

Are the French text extracts translations of the English ones?

Detecting pairwise dependence

… no doubt there is great pressure on provincial and municipal governments in relation to the issue of child care, but the reality is that there have been no cuts to child care funding from the federal government to the provinces. In fact, we have increased federal investments for early childhood development…

… il est évident que les ordres de gouvernements provinciaux et municipaux subissent de fortes pressions en ce qui concerne les services de garde, mais le gouvernement n'a pas réduit le financement qu'il verse aux provinces pour les services de garde. Au contraire, nous avons augmenté le financement fédéral pour le développement des jeunes enfants…

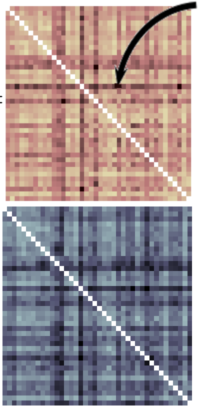


Detecting pairwise dependence

... no doubt there is great pressure on provincial and municipal governments in relation to the issue of child care, but the reality is that there have been no cuts to child care funding from the federal government to the provinces. In fact, we have increased federal investments for early childhood development...

$\longrightarrow$ $\mathbf{K}=$

$k(x_i, x_j)$



... il est évident que les ordres de gouvernements provinciaux et municipaux subissent de fortes pressions en ce qui concerne les services de garde, mais le gouvernement n'a pas réduit le financement qu'il verse aux provinces pour les services de garde. Au contraire, nous avons augmenté le financement fédéral pour le développement des jeunes enfants...

$\longrightarrow$ $\mathbf{L}=$

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

Kernel Embedding

- 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

- 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**
- 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**

Independence test via embeddings

- **Maximum Mean Discrepancy (MMD)** (Borgwardt et al, 2006; Gretton et al, 2007):
$MMD_k(P, Q) = \|\mu_k(P) - \mu_k(Q)\|_{\mathcal{H}_k}$
- **ISPD** kernels: μ_k injective on all signed measures and MMD_k metric (Sriperumbudur, 2010)
    - Gaussian, Laplacian, inverse multiquadratics, Matérn etc.

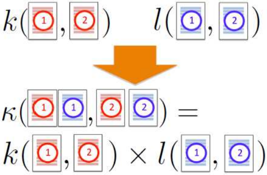
Independence test via embeddings

- **Maximum Mean Discrepancy (MMD)**

  (Borgwardt et al, 2006; Gretton et al, 2007):
  $MMD_k(P, Q) = \|\mu_k(P) - \mu_k(Q)\|_{\mathcal{H}_k}$
- **ISPD** kernels: μ_k injective on all signed measures and MMD_k metric (Sriperumbudur, 2010)
  - Gaussian, Laplacian, inverse multiquadratics, Matérn etc.
- **Hilbert-Schmidt Independence Criterion**

  Gretton et al (2005, 2008); Smola et al (2007):

  $$\left\|\mu_\kappa(\hat{P}_{XY}) - \mu_\kappa(\hat{P}_X \hat{P}_Y)\right\|^2_{\mathcal{H}_\kappa}$$



Independence test via embeddings

- **Maximum Mean Discrepancy (MMD)**

  (Borgwardt et al, 2006; Gretton et al, 2007):

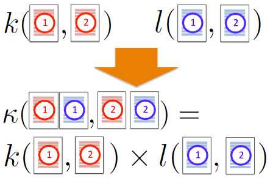
  $MMD_k(P,Q) = \|\mu_k(P) - \mu_k(Q)\|_{\mathcal{H}_k}$
- **ISPD** kernels: μ_k injective on all signed measures and MMD_k metric (Sriperumbudur, 2010)
  - Gaussian, Laplacian, inverse multiquadratics, Matérn etc.

- **Hilbert-Schmidt Independence Criterion**

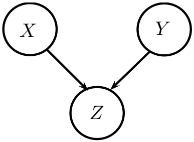
  Gretton et al (2005, 2008); Smola et al (2007):

  $$\left\|\mu_\kappa(\hat{P}_{XY}) - \mu_\kappa(\hat{P}_X\hat{P}_Y)\right\|^2_{\mathcal{H}_\kappa}$$
- $\mathrm{HSIC} = \frac{1}{n^2}\boxed{(HKH \circ HLH)_{++}}$

  Powerful independence tests that generalize dCov of Szekely et al (2007); DS et al (2013)



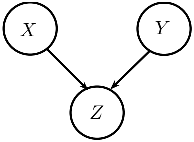
V-structure Discovery



Assume $X \perp\!\!\!\perp Y$ has been established (**first part**). V-structure can then be detected by:

- CI test: $\mathbf{H_0}: X \perp\!\!\!\perp Y | Z$ (Zhang et al 2011) or

V-structure Discovery



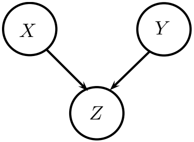
Assume $X \perp\!\!\!\perp Y$ has been established (**first part**). V-structure can then be detected by:

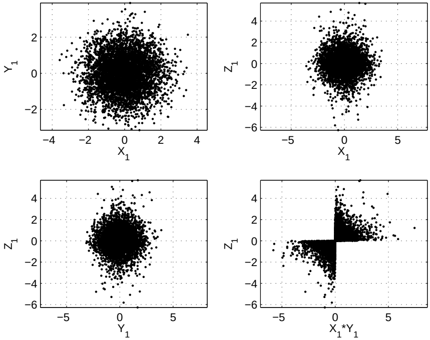
- CI test: $\mathbf{H_0} : X \perp\!\!\!\perp Y | Z$ (Zhang et al 2011) or
- Factorisation test: $\mathbf{H_0} : (X, Y) \perp\!\!\!\perp Z \vee (X, Z) \perp\!\!\!\perp Y \vee (Y, Z) \perp\!\!\!\perp X$ (multiple standard two-variable tests)
  - compute p-values for each of the marginal tests for $(Y,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

V-structure Discovery (2)

- How to detect V-structures with pairwise weak (or nonexistent) dependence?

- $X \perp\!\!\!\perp Y$, $Y \perp\!\!\!\perp Z$, $X \perp\!\!\!\perp Z$



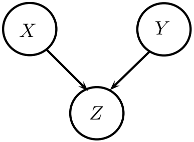


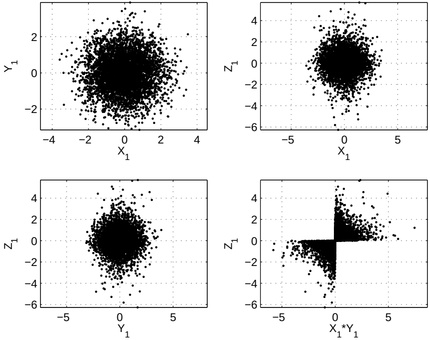
- $X_1, Y_1 \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$,
- $Z_1 | X_1, Y_1 \sim \mathrm{sign}(X_1 Y_1) Exp(\frac{1}{\sqrt{2}})$

V-structure Discovery (2)

- How to detect V-structures with pairwise weak (or nonexistent) dependence?

- $X \perp\!\!\!\perp Y$, $Y \perp\!\!\!\perp Z$, $X \perp\!\!\!\perp Z$





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

V-structure Discovery (3)

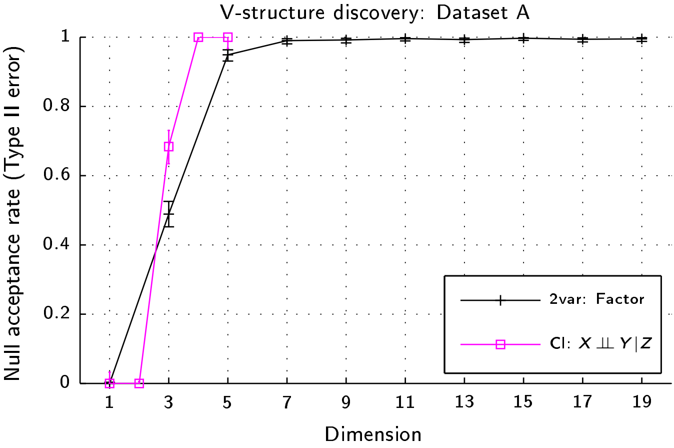


Figure: CI test for $X \perp\!\!\!\perp Y|Z$ from Zhang et al (2011), and a factorisation test with a HB correction, $n = 500$

Lancaster Interaction Measure

Definition (Bahadur (1961); Lancaster (1969))

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

Lancaster Interaction Measure

Definition (Bahadur (1961); Lancaster (1969))

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

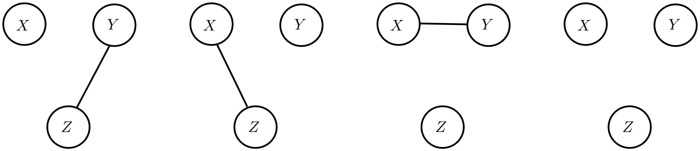
- $D = 2:$ $\quad \Delta_L P = P_{XY} - P_X P_Y$

Lancaster Interaction Measure

Definition (Bahadur (1961); Lancaster (1969))

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

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



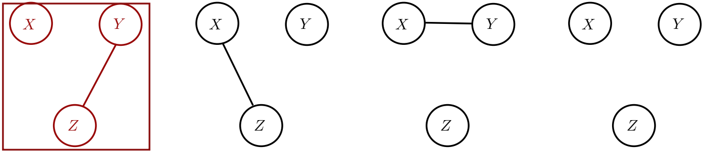
$$\Delta_L P = P_{XYZ} \quad -P_X P_{YZ} \quad -P_{XZ} P_Y \quad -P_{XY} P_Z \quad +2 P_X P_Y P_Z$$

Lancaster Interaction Measure

Definition (Bahadur (1961); Lancaster (1969))

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

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



$\Delta_L P = 0$

~~P_{XYZ}~~ ~~$-P_X P_{YZ}$~~ ~~$-P_{XZ} P_Y$~~ ~~$-P_{XY} P_Z$~~ ~~$+2 P_X P_Y P_Z$~~

A Test using Lancaster Measure

- Construct a test by estimating $\|\mu_\kappa(\Delta_L P)\|^2_{\mathcal{H}_\kappa}$, where $\kappa = k \otimes l \otimes m$:

$$\|\mu_\kappa(P_{XYZ} - P_{XY}P_Z - \cdots)\|^2_{\mathcal{H}_\kappa} =$$
$$\langle \mu_\kappa P_{XYZ}, \mu_\kappa P_{XYZ} \rangle_{k\otimes l\otimes m} - 2\langle \mu_\kappa P_{XYZ}, \mu_\kappa P_{XY}P_Z \rangle_{k\otimes l\otimes m} \cdots$$

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}	$(K \circ L \circ M)_{++}$	$((K \circ L)\, M)_{++}$	$((K \circ M)\, L)_{++}$	$((M \circ L)\, K)_{++}$	$tr(K_+ \circ L_+ \circ M_+)$
$P_{XY}P_Z$		$(K \circ L)_{++}\, M_{++}$	$(MKL)_{++}$	$(KLM)_{++}$	$(KL)_{++}M_{++}$
$P_{XZ}P_Y$			$(K \circ M)_{++}\, L_{++}$	$(KML)_{++}$	$(KM)_{++}L_{++}$
$P_{YZ}P_X$				$(L \circ M)_{++}\, K_{++}$	$(LM)_{++}K_{++}$
$P_X P_Y P_Z$					$K_{++}L_{++}M_{++}$

Table: V-statistic estimators of $\langle \mu_\kappa \nu, \mu_\kappa \nu' \rangle_{k \otimes l \otimes m}$ in the three-variable case

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}	$(K \circ L \circ M)_{++}$	$((K \circ L)\, M)_{++}$	$((K \circ M)\, L)_{++}$	$((M \circ L)\, K)_{++}$	$tr(K_+ \circ L_+ \circ M_+)$
$P_{XY}P_Z$		$(K \circ L)_{++}\, M_{++}$	$(MKL)_{++}$	$(KLM)_{++}$	$(KL)_{++}M_{++}$
$P_{XZ}P_Y$			$(K \circ M)_{++}\, L_{++}$	$(KML)_{++}$	$(KM)_{++}L_{++}$
$P_{YZ}P_X$				$(L \circ M)_{++}\, K_{++}$	$(LM)_{++}K_{++}$
$P_X P_Y P_Z$					$K_{++}L_{++}M_{++}$

Table: V-statistic estimators of $\langle \mu_\kappa \nu, \mu_\kappa \nu' \rangle_{k \otimes l \otimes m}$ in the three-variable case

Proposition (Lancaster interaction statistic)

$$\|\mu_\kappa(\Delta_L P)\|^2_{\mathcal{H}_\kappa} = \frac{1}{n^2}\boxed{(HKH \circ HLH \circ HMH)_{++}}.$$

Empirical joint central moment in the feature space

Example A: factorization tests

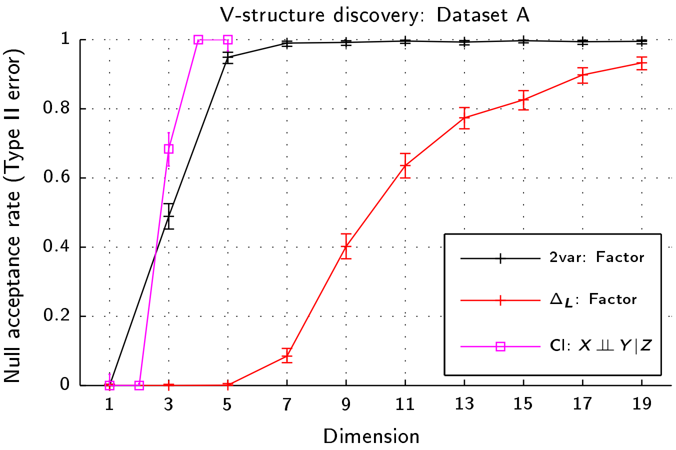


Figure: Factorization hypothesis: Lancaster statistic vs. a two-variable based test (both with HB correction); Test for $X \perp\!\!\!\perp Y|Z$ from Zhang et al (2011), $n = 500$

Example B: Joint dependence can be easier to detect

- A triplet of random vectors (X, Y, Z) on $\mathbb{R}^p \times \mathbb{R}^p \times \mathbb{R}^p$, with $X, Y \overset{i.i.d.}{\sim} \mathcal{N}(0, I_p)$, $Z_{2:p} \sim \mathcal{N}(0, I_{p-1})$, and

$$Z_1 = \begin{cases} X_1^2 + \epsilon, & w.p.\ 1/3, \\ Y_1^2 + \epsilon, & w.p.\ 1/3, \\ X_1 Y_1 + \epsilon, & w.p.\ 1/3. \end{cases}$$

  where $\epsilon \sim \mathcal{N}(0, 0.1^2)$.
- dependence of Z on pair (X, Y) is stronger than on X and Y individually

Example B: factorization tests

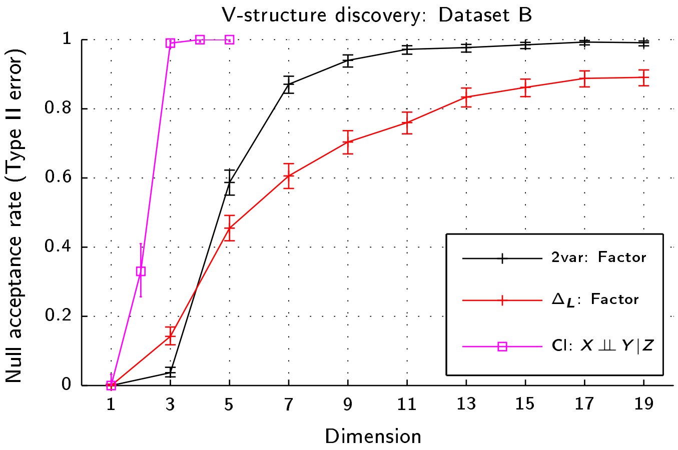


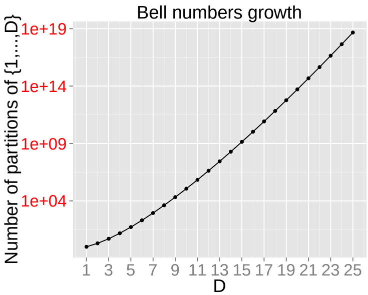
Figure: Factorization hypothesis: Lancaster statistic vs. a two-variable based test; Test for $X \perp\!\!\!\perp Y|Z$ from Zhang et al (2011), $n = 500$

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., $J_{13|2|4} P = P_{X_1 X_3} P_{X_2} P_{X_4}$.

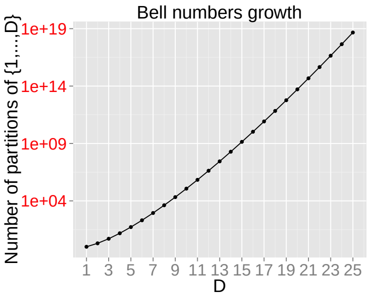


Interaction for $D \geq 4$

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

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

  - For a partition π, J_π associates to the joint the corresponding factorization, e.g., $J_{13|2|4}P = P_{X_1X_3}P_{X_2}P_{X_4}$.



joint central moments (Lancaster interaction)

vs.

joint cumulants (Streitberg interaction)

Summary

- A nonparametric test for three-variable interaction and for total independence, using embeddings of signed measures into RKHSs

Summary

- A nonparametric test for three-variable interaction and for total independence, using embeddings of signed measures into RKHSs
- Test statistics are simple and easy to compute - corresponding permutation tests significantly outperform standard two-variable-based tests on V-structures with weak pairwise interactions

Summary

- A nonparametric test for three-variable interaction and for total independence, using embeddings of signed measures into RKHSs
- Test statistics are simple and easy to compute - corresponding permutation tests significantly outperform standard two-variable-based tests on V-structures with weak pairwise interactions
- All forms of Lancaster three-variable interaction can be detected for a large family of reproducing kernels (**ISPD**)

References

- B. Streitberg, **Lancaster interactions revisited**. *Annals of Statistics* 18(4): 1878–1885, 1990.
- A. Kankainen. *Consistent Testing of Total Independence Based on the Empirical Characteristic Function*. PhD thesis, University of Jyväskylä, 1995.
- A. Gretton, K. Fukumizu, C.-H. Teo, L. Song, B. Schölkopf and A. Smola. **A kernel statistical test of independence**. in *Advances in Neural Information Processing Systems* 20: 585–592, MIT Press, 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.
- D. Sejdinovic, B. Sriperumbudur, A. Gretton and K. Fukumizu, **Equivalence of distance-based and RKHS-based statistics in hypothesis testing**. *Annals of Statistics* 41(5): 2263-2291, 2013.

Total independence test

- Total independence test:
  $\mathbf{H_0}: P_{XYZ} = P_X P_Y P_Z$ vs. $\mathbf{H}_1: P_{XYZ} \neq P_X P_Y P_Z$

Total independence test

- Total independence test:
  $\mathbf{H_0}: P_{XYZ} = P_X P_Y P_Z$ vs. $\mathbf{H}_1: P_{XYZ} \neq P_X P_Y P_Z$
- For $(X_1, \ldots, X_D) \sim P_{\mathbf{X}}$, and $\kappa = \bigotimes_{i=1}^{D} k^{(i)}$:

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

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

Example B: total independence tests

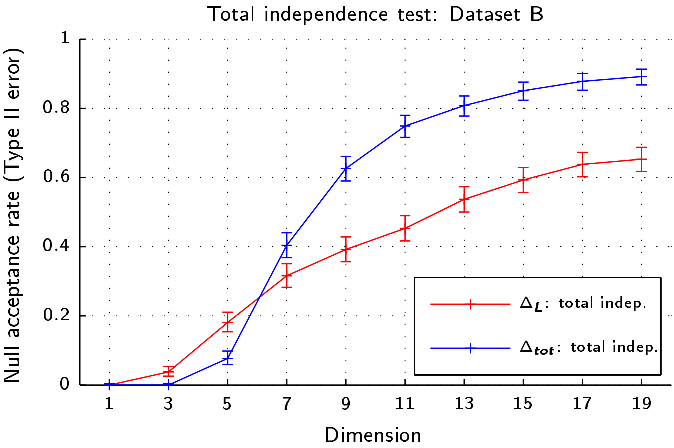


Figure: Total independence: $\Delta_{tot}\hat{P}$ vs. $\Delta_L\hat{P}$, $n = 500$