A Kernel Test for Three Variable Interactions

Dino Sejdinovic\textsuperscript{1}, Arthur Gretton\textsuperscript{1}, Wicher Bergsma\textsuperscript{2}

\textsuperscript{1}Gatsby Unit, CSML, University College London
\textsuperscript{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?

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

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

![Diagram](https://example.com/diagram.png)
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.

**Y_1:** Honorables sénateurs, ma question s’adresse au leader du gouvernement au Sénat et concerne l’aide financière 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...

... 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...

\[ k(x_i, x_j) \]

\[ \langle HKH, HLH \rangle = (HKH \circ HLH)_{++} \]

\[ H = I - \frac{1}{n} \mathbf{1} \mathbf{1}^\top \] (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**: \( \mu_k \) 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**: \( \mu_k \) injective on all signed measures and \( MMD_k \) metric
  
  (Sriperumbudur, 2010)

  - Gaussian, Laplacian, inverse multiquadratics, Matérn etc.

- **Hilbert-Schmidt Independence Criterion**


  \[ \|\mu_\kappa(\hat{P}_{XY}) - \mu_\kappa(\hat{P}_X \hat{P}_Y)\|_{\mathcal{H}_\kappa}^2 \]

\[
\begin{align*}
 k(1, 2) & \quad l(1, 2) \\
 k(1, 1, 2, 2) & = \\
 k(1, 2) \times l(1, 2)
\end{align*}
\]
Independence test via embeddings

- **Maximum Mean Discrepancy (MMD)**
  
  \[ MMD_k(P, Q) = \| \mu_k(P) - \mu_k(Q) \|_{\mathcal{H}_k} \]

- **ISPD kernels**: \( \mu_k \) injective on all signed measures and MMD\(_k\) metric
  
  - Gaussian, Laplacian, inverse multiquadratics, Matérn etc.

- **Hilbert-Schmidt Independence Criterion**
  
  \[ \text{HSIC} = \frac{1}{n^2} \left( (HKH \circ HLL)_{++} \right) \]

  Powerful independence tests that generalize dCov of Szekely et al (2007); DS et al (2013)
Assume $X \perp Y$ has been established (first part). V-structure can then be detected by:

- **CI test:** $H_0: X \perp Y | Z$ (Zhang et al 2011) or
V-structure Discovery

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

- Cl test: $H_0 : X \perp Y | Z$ (Zhang et al 2011) or

- Factorisation test: $H_0 : (X, Y) \perp Z \lor (X, Z) \perp Y \lor (Y, Z) \perp X$
  (multiple standard two-variable tests)
    - compute $p$-values for each of the marginal tests for $(Y, Z) \perp X$, $(X, Z) \perp Y$, or $(X, Y) \perp Z$
    - apply Holm-Bonferroni (HB) sequentially rejective correction (Holm 1979)
V-structure Discovery (2)

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

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

\[
X, Y_1 \overset{i.i.d.}{\sim} \mathcal{N}(0,1),
\]

\[
Z_1 \mid X_1, Y_1 \sim \text{sign}(X_1 Y_1) \text{Exp}(\frac{1}{\sqrt{2}})
\]
V-structure Discovery (2)

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

- \( X \perp Y, Y \perp Z, X \perp Z \)

\[ \begin{align*}
  &X_1, Y_1 \overset{i.i.d.}{\sim} \mathcal{N}(0, 1), \\
  &Z_1 \mid X_1, Y_1 \sim \\
  &\text{sign}(X_1 Y_1) \Exp\left(\frac{1}{\sqrt{2}}\right) \\
  &X_{2:p}, Y_{2:p}, Z_{2:p} \overset{i.i.d.}{\sim} \mathcal{N}(0, I_{p-1})
\end{align*} \]
V-structure Discovery (3)

Figure: Cl test for $X \perp Y | Z$ 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 \(\Delta P\) 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 \(\Delta P\) that vanishes whenever \(P\) can be factorised in a non-trivial way as a product of its (possibly multivariate) marginal distributions.

- \(D = 2:\) \(\Delta_L P = P_{XY} - P_X P_Y\)
**Lancaster Interaction Measure**

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

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

- **\(D = 2\):** \(\Delta_L P = P_{XY} - P_X P_Y\)
- **\(D = 3\):**
  
  \[
  \Delta_L P = P_{XYZ} - P_X P_{YZ} - P_{XZ} P_Y - P_{XY} P_Z + 2 P_X P_Y P_Z
  \]
Lancaster Interaction Measure

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

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

- \(D = 2:\) \[\Delta_L P = P_{XY} - P_X P_Y\]
- \(D = 3:\)
  - \(\Delta_L P = 0\)
  - \(P_{XYZ} = -P_{XZ} P_Y\)
  - \(-P_{XZ} P_Y + P_{XY} P_Z + 2P_X P_Y P_Z\)
Construct a test by estimating $\|\mu_\kappa(\Delta L P)\|_{\mathcal{H}_\kappa}^2$, where $\kappa = k \otimes l \otimes m$:

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

<table>
<thead>
<tr>
<th>$\nu \setminus \nu'$</th>
<th>$P_{XYZ}$</th>
<th>$P_{XY}P_Z$</th>
<th>$P_{XZ}P_Y$</th>
<th>$P_{YZ}P_X$</th>
<th>$P_XP_YP_Z$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_{XYZ}$</td>
<td>$(K \circ L \circ M)_{++}$</td>
<td>$((K \circ L) M)_{++}$</td>
<td>$((K \circ M) L)_{++}$</td>
<td>$((M \circ L) K)_{++}$</td>
<td>$tr(K_+ \circ L_+ \circ M_+)$</td>
</tr>
<tr>
<td>$P_{XY}P_Z$</td>
<td>$(K \circ L)<em>{++} M</em>{++}$</td>
<td>$(MKL)_{++}$</td>
<td>$(KLM)_{++}$</td>
<td>$(KL)<em>{++} M</em>{++}$</td>
<td></td>
</tr>
<tr>
<td>$P_{XZ}P_Y$</td>
<td>$(K \circ M)<em>{++} L</em>{++}$</td>
<td>$K_{++} L_{++}$</td>
<td>$(KM)<em>{++} L</em>{++}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$P_{YZ}P_X$</td>
<td>$L \circ M)<em>{++} K</em>{++}$</td>
<td>$(LM)<em>{++} K</em>{++}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$P_XP_YP_Z$</td>
<td></td>
<td></td>
<td></td>
<td>$K_{++} L_{++} M_{++}$</td>
<td></td>
</tr>
</tbody>
</table>

**Table:** $V$-statistic estimators of $\langle \mu_K \nu, \mu_K \nu' \rangle_{k \otimes l \otimes m}$ in the three-variable case.
## Inner Product Estimators

<table>
<thead>
<tr>
<th>$\nu \backslash \nu'$</th>
<th>$P_{XYZ}$</th>
<th>$P_{XY}P_Z$</th>
<th>$P_{XZ}P_Y$</th>
<th>$P_{YZ}P_X$</th>
<th>$P_XP_YP_Z$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_{XYZ}$</td>
<td>$(K \circ L \circ M)_{++}$</td>
<td>$((K \circ L) M)_{++}$</td>
<td>$((K \circ M) L)_{++}$</td>
<td>$((M \circ L) K)_{++}$</td>
<td>$\text{tr}(K_+ \circ L_+ \circ M_+)$</td>
</tr>
<tr>
<td>$P_{XY}P_Z$</td>
<td>$(K \circ L)<em>{++} M</em>{++}$</td>
<td>$(MKL)_{++}$</td>
<td>$(KLM)_{++}$</td>
<td>$(KL)<em>{++} M</em>{++}$</td>
<td>$K_{++}$</td>
</tr>
<tr>
<td>$P_{XZ}P_Y$</td>
<td>$(K \circ M)<em>{++} L</em>{++}$</td>
<td>$(KML)_{++}$</td>
<td>$(KM)<em>{++} L</em>{++}$</td>
<td>$L_{++}$</td>
<td></td>
</tr>
<tr>
<td>$P_{YZ}P_X$</td>
<td>$(L \circ M)<em>{++} K</em>{++}$</td>
<td>$(LM)<em>{++} K</em>{++}$</td>
<td>$K_{++}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$P_XP_YP_Z$</td>
<td></td>
<td></td>
<td>$K_{++} L_{++} M_{++}$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**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)

$$\left\| \mu_\kappa (\Delta LP) \right\|_{H_\kappa}^2 = \frac{1}{n^2} \left( HKH \circ HLL \circ HMM \right)_{++}.$$  

**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 $X \perp Y|Z$ 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

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

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 $X \perp Y \mid Z$ 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 $\pi$, $J_{\pi}$ 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):

  \[ \Delta_s P = \sum_{\pi} (-1)^{|\pi|-1} (|\pi| - 1)! J_\pi P \]

- For a partition $\pi$, $J_\pi$ associates to the joint the corresponding factorization, e.g., $J_{13|2|4} P = P_{x_1 x_3} P_{x_2} P_{x_4}$.

**joint central moments** (Lancaster interaction)

vs.

**joint cumulants** (Streitberg interaction)

Bell numbers growth
Summary

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


Total independence test

- **Total independence test:**
  \[
  H_0 : P_{XYZ} = P_X P_Y P_Z \text{ vs. } H_1 : P_{XYZ} \neq P_X P_Y P_Z
  \]
Total independence test

- Total independence test:
  \[ H_0 : P_{XYZ} = P_X P_Y P_Z \text{ vs. } H_1 : P_{XYZ} \neq P_X P_Y P_Z \]

- For \((X_1, \ldots, X_D) \sim P_X\), and \(\kappa = \bigotimes_{i=1}^D k(i)\):

\[
\begin{align*}
\mathbb{E}_{\kappa} \left( \left\| \hat{P}_X - \prod_{i=1}^D \hat{P}_{X_i} \right\|_2^2 \right)_{\Delta_{\text{tot}} \hat{P}} &=
\frac{1}{n^2} \sum_{a=1}^n \sum_{b=1}^n \prod_{i=1}^D K_{ab}^{(i)} - \frac{2}{n^{D+1}} \sum_{a=1}^n \prod_{i=1}^D \sum_{b=1}^n K_{ab}^{(i)} \\
&+ \frac{1}{n^{2D}} \prod_{i=1}^D \sum_{a=1}^n \sum_{b=1}^n K_{ab}^{(i)}.
\end{align*}
\]

- 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. $\Delta_L \hat{P}$, $n = 500$