Aggregated tests of independence based on HSIC measures

Anouar Meynaoui, Mélisande Albert, Béatrice Laurent, Amandine Marrel

INSA de Toulouse
Institut de Mathématiques de Toulouse, France
CEA, DEN, DER, France

Aggregated tests of independence based on HSIC measures (part 2)

European Meeting of Statisticians, 2019

Anouar Meynaoui, Mélisande Albert, Béatrice Laurent, Amandine Marrel

Outline

Introduction

- We recall that we **study the independence of two real random vectors** $X = \left(X^{(1)}, \dots, X^{(p)}\right)$ and $Y = \left(Y^{(1)}, \dots, Y^{(q)}\right)$ with marginal densities resp. denoted f_1 and f_2 and joint density f.
- We recall that we have an i.i.d. sample $\mathbb{Z}_n = (X_i, Y_i)_{1 \leq i \leq n}$ of (X, Y).
- We rely on HSIC-based independence tests with **Gaussian kernels** k_λ and l_μ resp. associated to X and Y.
- **In the previous talk,** we first proposed for each couple of values (λ, μ) a **theoretical** HSIC test of independence of level α in $(0, 1)$, followed by a **non-asymptotic permutation-based** test, of the same level α.
- The **power of the permuted test** is shown to be approximately the same as **theoretical power** if enough permutations are used.

Introduction

- When $f - f_1 \otimes f_2$ belongs to a **Sobolev ball** with regularity δ in $(0, 2]$, **sharp upper bounds** of the uniform separation rate w.r.t. the values of λ and μ are provided.
- The HSIC test with the **optimal upper bound** is shown to be **minimax** over Sobolev balls.
- This optimal test is **not adaptive,** since it depends on the regularity δ.
- **In this talk,** we provide an **adaptive procedure of testing independence** which doesn't depend on the regularity δ.
- This procedure is based on the **aggregation** of a collection of HSIC-tests with a collection of different bandwidths λ and μ.
- Numerical studies to **assess the performance** of the procedure and to **compare methodological choices** are then provided.

The aggregated testing procedure

- Single HSIC-based test leads to the **question of the choice of kernel bandwidths** λ and μ. **Heuristic choices are adopted in practice**, with no theoretical justifications.
- We propose here an **aggregated testing procedure** combining a **collection** of single tests based on different bandwidths.
- We consider a finite or countable collection $\Lambda \times U$ of bandwidths in $(0, +\infty)^p \times (0, +\infty)^q$ and a collection of positive weights $\{\omega_{\lambda,\mu} \ / \ (\lambda, \mu) \in \Lambda \times U\}$ such that $\sum_{(\lambda,\mu)\in\Lambda\times U} e^{-\omega_{\lambda,\mu}} \leq 1$.

The aggregated testing procedure

- For a given $\alpha \in (0,1)$, we define the aggregated test Δ_α which rejects $(\mathcal{H}_0)$ if there is at least one $(\lambda,\mu) \in \Lambda \times U$ such that

$$\widehat{\mathrm{HSIC}}_{\lambda,\mu} \; > \; q^{\lambda,\mu}_{1-u_\alpha e^{-\omega_{\lambda,\mu}}},$$

where u_α is the less conservative value such that the test is of level α, and is defined by

$$u_\alpha = \sup\left\{ u > 0 \; ; \; P_{f_1 \otimes f_2}\left(\sup_{(\lambda,\mu)\in\Lambda\times U} \left(\widehat{\mathrm{HSIC}}_{\lambda,\mu} - q^{\lambda,\mu}_{1-ue^{-\omega_{\lambda,\mu}}} \right) > 0 \right) \leq \alpha \right\}.$$

- The test function Δ_α associated to this aggregated test, takes values in $\{0,1\}$ and is defined by

$$\Delta_\alpha = 1 \quad \Longleftrightarrow \quad \sup_{(\lambda,\mu)\in\Lambda\times U} \left(\widehat{\mathrm{HSIC}}_{\lambda,\mu} - q^{\lambda,\mu}_{1-u_\alpha e^{-\omega_{\lambda,\mu}}} \right) > 0.$$

Oracle type conditions for the second kind error

- The aggregated testing procedure Δ_α is of **level** α.
- The **second kind** error of the aggregated testing procedure Δ_α verifies the inequality

$$\mathbb{P}_f\left(\Delta_\alpha = 0\right) \leq \inf_{(\lambda,\mu)\in\Lambda\times U}\left\{\mathbb{P}_f\left(\Delta^{\lambda,\mu}_{\alpha e^{-\omega_{\lambda,\mu}}} = 0\right)\right\},$$

where $\Delta^{\lambda,\mu}_{\alpha e^{-\omega_{\lambda,\mu}}}$ is the single test of level $\alpha e^{-\omega_{\lambda,\mu}}$ associated to the bandwidths (λ, μ)

- The aggregated testing procedure has a **second kind at most equal to** β, if there exists at least one $(\lambda, \mu) \in \Lambda \times U$ such that the test $\Delta^{\lambda,\mu}_{\alpha e^{-\omega_{\lambda,\mu}}}$ has a probability of second kind error at most equal to β.

Oracle type conditions for the second kind error

Theorem

Let $\alpha, \beta \in (0,1)$, $\left\{(k_\lambda, l_\mu) \;/\; (\lambda,\mu) \in \Lambda \times U\right\}$ a collection of Gaussian kernels and $\left\{\omega_{\lambda,\mu} \;/\; (\lambda,\mu) \in \Lambda \times U\right\}$ a collection of positive weights, such that $\sum_{(\lambda,\mu)\in\Lambda\times U} e^{-\omega_{\lambda,\mu}} \leq 1$.

We assume that f, f_1 and f_2 are bounded. We also assume that all bandwidths (λ,μ) in $\Lambda \times U$ verify the following conditions

$$\max\left(\lambda_1 \ldots \lambda_p\ ,\ \mu_1 \ldots \mu_q\right) < 1 \text{ and } n\sqrt{\lambda_1 \ldots \lambda_p \mu_1 \ldots \mu_q} > \log\left(\frac{1}{\alpha}\right) > 1.$$

Then, the uniform separation rate $\rho\left(\Delta_\alpha, \mathcal{S}^\delta_{p+q}(R), \beta\right)$, where $\delta \in (0,2]$ and $R > 0$ can be upper bounded as follows

Oracle type conditions for the second kind error

$$\left[\rho\left(\Delta_\alpha, \mathcal{S}^{\delta}_{p+q}(R), \beta\right)\right]^2 \leq C\left(M_f, p, q, \beta, \delta\right) \inf_{(\lambda,\mu)\in\Lambda\times U} \left\{ \frac{1}{n\sqrt{\lambda_1\ldots\lambda_p\mu_1\ldots\mu_q}} \left(\log\left(\frac{1}{\alpha}\right) + \omega_{\lambda,\mu}\right) + \left[\sum_{i=1}^{p} \lambda_i^{2\delta} + \sum_{j=1}^{q} \mu_j^{2\delta}\right] \right\}$$

where $M_f = \max\left(\|f\|_\infty, \|f_1\|_\infty, \|f_2\|_\infty\right)$ and $C\left(M_f, p, q, \beta, \delta\right)$ is a positive constant depending only on its arguments.

- This theorem gives an **oracle type condition** of the uniform separation rate. Indeed, without knowing the regularity of $f - f_1 \otimes f_2$, we prove that the uniform separation rate of Δ_α is of the same order as the smallest uniform separation rate over $(\lambda, \mu) \in \Lambda \times U$, up to $\omega_{\lambda,\mu}$.

Adaptive procedure of testing independence

- We consider the bandwidth collections Λ and U defined by

$$\Lambda = \{(2^{-m_{1,1}}, \ldots, 2^{-m_{1,p}}) \ ; \ (m_{1,1}, \ldots, m_{1,p}) \in (\mathbb{N}^*)^p\}, \quad (1)$$

$$U = \{(2^{-m_{2,1}}, \ldots, 2^{-m_{2,q}}) \ ; \ (m_{2,1}, \ldots, m_{2,q}) \in (\mathbb{N}^*)^q\}. \quad (2)$$

- We associate to every $\lambda = (2^{-m_{1,1}}, \ldots, 2^{-m_{1,p}})$ in Λ and $\mu = (2^{-m_{2,1}}, \ldots, 2^{-m_{2,q}})$ in U the positive weights

$$\omega_{\lambda,\mu} = 2\sum_{i=1}^{p} \log\left(m_{1,i} \times \frac{\pi}{\sqrt{6}}\right) + 2\sum_{j=1}^{q} \log\left(m_{2,j} \times \frac{\pi}{\sqrt{6}}\right), \quad (3)$$

so that $\sum_{(\lambda,\mu)\in\Lambda\times U} e^{-\omega_{\lambda,\mu}} = 1$.

Adaptive procedure of testing independence

Corollary

Assuming that $\log\log(n) > 1$, $\alpha, \beta \in (0,1)$ *and* Δ_α *the aggregated testing procedure, with the particular choice of* Λ, U *and the weights* $(\omega_{\lambda,\mu})_{(\lambda,\mu)\in\Lambda\times U}$ *defined in* (1), (2) *and* (3). *Then, the uniform separation rate* $\rho\left(\Delta_\alpha, \mathcal{S}^{\delta}_{p+q}(R), \beta\right)$ *of the aggregated test* Δ_α *over Sobolev spaces where* δ *in* $(0,2]$, *can be upper bounded as follows*

$$\rho\left(\Delta_\alpha, \mathcal{S}^{\delta}_{p+q}(R), \beta\right) \leq C\left(M_f, p, q, \alpha, \beta, \delta\right)\left(\frac{\log\log(n)}{n}\right)^{\frac{2\delta}{4\delta+(p+q)}},$$

where $M_f = \max\left(\|f\|_\infty, \|f_1\|_\infty, \|f_2\|_\infty\right)$.

- The rate of the aggregation procedure over the classes of Sobolev balls is in the **same order of the smallest rate of single tests,** up to a *loglog*(n) factor. This combined with the result on the **lower bound over Sobolev** shows that the **aggregated test is adaptive** over these regularity classes.

Implementation of the aggregated procedure

- The collections Λ and U are finite in practice.
- The correction u_α defined as

$$u_\alpha = \sup\left\{ u > 0\,;\; P_{f_1 \otimes f_2}\left(\sup_{(\lambda,\mu)\in\Lambda\times U} \left(\widehat{\mathrm{HSIC}}_{\lambda,\mu} - q^{\lambda,\mu}_{1-ue^{-\omega_{\lambda,\mu}}} \right) > 0 \right) \leq \alpha \right\}.$$

can be approached by a permutation method with Monte Carlo approximation, as done in Albert et al., 2015.

- **To compute the quantiles** $\hat{q}^{\lambda,\mu}_{1-ue^{-\omega_{\lambda,\mu}}}$, we generate uniformly $\mathbf{B_1}$ independent random permutations $\tau_1, \ldots, \tau_{B_1}$, independent of $\mathbb{Z}_n$. We then compute for each $(\lambda,\mu) \in \Lambda \times U$ and each $u > 0$ the permuted quantile **with Monte Carlo approximation** $\hat{q}^{\lambda,\mu}_{1-ue^{-\omega_{\lambda,\mu}}}$.

Implementation of the aggregated procedure

- **To compute the probability** $P_{f_1 \otimes f_2}$, we generate uniformly $\mathbf{B_2}$ independent random permutations $\kappa_1, ..., \kappa_{B_2}$, independent of $\mathbb{Z}_n$. Denote for all permutation κ_b, the corresponding permuted statistic

$$\widehat{H}^{\kappa_b}_{\lambda,\mu} = \widehat{\mathrm{HSIC}}_{\lambda,\mu}\left(\mathbb{Z}_n^{\kappa_b}\right).$$

- Then, the correction u_α is approached by

$$\hat{u}_\alpha = \sup \left\{ u > 0 \; ; \; \frac{1}{B_2} \sum_{b=1}^{B_2} \mathbb{1}_{\max_{(\lambda,\mu)\in\Lambda\times U}\left\{ \widehat{H}^{\kappa_b}_{\lambda,\mu} - \hat{q}^{\lambda,\mu}_{1-ue^{-\omega_{\lambda,\mu}}} \right\} > 0} \leq \alpha \right\}. \tag{4}$$

- The supremum in Equation (4) is estimated by **dichotomy**.
- **Simulation result:** the powers of the implemented and the theoretical procedures are **approximately the same if enough permutation** are used.

Analytical examples

Dependence forms from Berrett and Samworth., 2017:

(i). Defining the joint density $f_l, l = 1, \ldots, 10$ of (X, Y) on $[-\pi, \pi]$ by

$$f_l(x, y) = \frac{1}{4\pi^2} \left\{1 + \sin(lx) \sin(ly)\right\}.$$

(ii). Considering X and Y as

$$X = L \cos \Theta + \frac{\varepsilon_1}{4}, \quad Y = L \sin \Theta + \frac{\varepsilon_2}{4},$$

where L, Θ, ε_1 and ε_2 are independent, with $L \sim \mathcal{U}\{1, \ldots, l\}$ for $l = 1, \ldots, 10$, $\Theta \sim \mathcal{U}[0, 2\pi]$ and $\varepsilon_1, \varepsilon_2 \sim \mathcal{N}(0, 1)$.

(iii). Defining $X \sim \mathcal{U}[-1, 1]$. For a given $\rho = 0.1, 0.2, \ldots, 1$, Y is defined as $Y = |X|^\rho \varepsilon$, where $\varepsilon \sim \mathcal{N}(0, 1)$ independent with X.

Collection of bandwidths

- **Choice of collections Λ and U:** recommendation of dyadic collection, multiple and dividers by powers of 2 of the X and Y standard deviations (respectively noted s and s' in Figure 1).

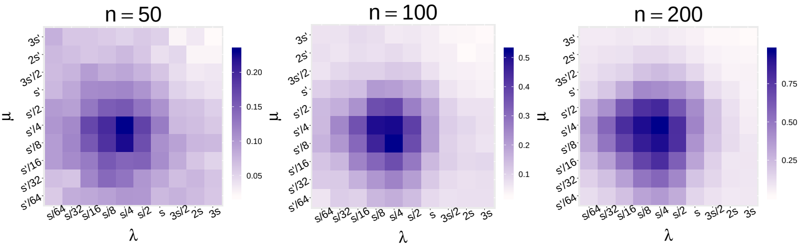

FIGURE 1: Analytical example (ii), $l = 2$. Power map of single HSIC test w.r.t. to kernel widths λ and μ respectively associated to X and Y, for sample sizes $n = 50, 100$ and 200.

Weights associated to the collection

- **Choice of weights:** comparison of **uniform** and **exponential decreasing** weights in Figure 2.

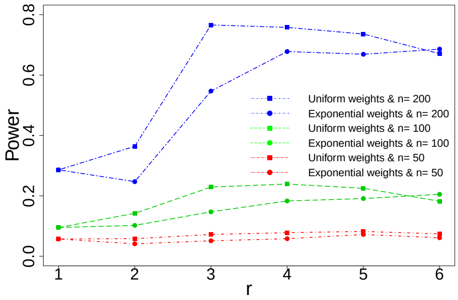

FIGURE 2: Analytical example (ii), $l = 2$. Power of aggregated procedures with uniform and exponential weights, w.r.t. the number r of aggregated widths in each direction, for sample sizes $n = 50, 100$ and 200.

Comparison with other independence tests

- **Comparison with** Single HSIC using the permutation method (De Lozzo et Marrel, 2016 ; Meynaoui et al., 2019) and the Mutual Information Test (MINT, Berrett et Samworth, 2017).

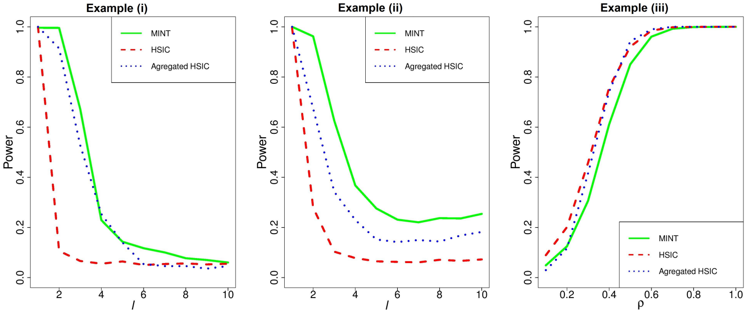

FIGURE 3: Power curves of MINT, single HSIC test and aggregated procedure for the mechanisms of dependence (i), (ii) and (iii).

Conclusion and Prospect

- Proposition of a test procedure based on **aggregating single HSIC tests** with different choices of bandwidths.
- Procedure **adaptive over Sobolev balls,** i.e. achieving the optimal uniform separation rate and does not depend on any regularity parameter.
- Encouraging results (on terms of test power) on some analytical examples.

Some possible improvements:

- **Extend** the aggregation procedure to **other characteristic kernels** and **other types of random variables** (e.g. discrete variables).
- **Extend** the aggregation procedure to **other types of experimental designs** such as Quasi-Monte Carlo and Space Filling Designs.
- A confrontation of the methodology to a **real data case** is in progress. Data stem from an **industrial case simulating a severe nuclear reactor accidental scenario**.

References

- Albert, M., Bouret, Y., Fromont, M., Reynaud-Bouret, P., et al. (2015). Bootstrap and permutation tests of independence for point processes. *The Annals of Statistics, 43(6):2537–2564.*
- Berrett, T. B. and Samworth, R. J. (2017). Nonparametric independence testing via mutual information. *arXiv preprint arXiv :1711.06642.*
- De Lozzo, M. and Marrel, A. (2016b). New improvements in the use of dependence measures for sensitivity analysis and screening. *Journal of Statistical Computation and Simulation, 86(15) :3038–3058.*
- Gretton, A., Bousquet, O., Smola, A.and Scholkopf, B., Measuring statistical dependence with Hilbert-Schmidt norms, *ALT, 2005.*
- Meynaoui, A., Albert, M., Laurent, B., and Marrel, A. (2019). Aggregated test of independence based on hsic measures. *arXiv preprint arXiv :1902.06441.*

Acknowledgements.

The authors would like to thank the Innovation and Industrial Nuclear Support Division of CEA for funding this CEA PhD work performed in the frame of codes development for Generation IV nuclear reactor safety studies.