An Asymptotic Test for Conditional Independence using Analytic Kernel Embeddings

FACEBOOK AI — TECHNION Israel Institute of Technology — ENSAE IP PARIS — Dauphine Université Paris PSL

Meyer SCETBON, Laurent MEUNIER*, Yaniv ROMANO*

Overview

Problem: How to infer from data conditional dependencies between random variables?

Contributions:

- We design a simple and consistent kernel-based conditional independence test using a randomized version of the ℓ_p distance between analytic kernel mean embeddings.
- We characterize the conditional independence between random variables using this distance, derive a first oracle estimate of it and obtain its asymptotic null distribution.
- We then propose an approximation of the oracle statistic using regularized least-squares estimators and show that it has the same asymptotic distribution under some mild assumptions.
- In order to obtain a simple null asymptotic distribution, we consider also a normalized version of our tractable test statistic and show that it converges towards a standard normal distribution under the null hypothesis.
- Finally we show on various experiments that our test outperforms other SoTA tests as it is the only one able to control the Type-I error and obtain high power.

ℓ_p Distance between MEs

Definition

Let k be a positive definite, continuous, bounded and **analytic** kernel on $\mathbb{R}^d$, P, Q two distributions on $\mathbb{R}^d$, denote respectively $\mu_{P,k}$ and $\mu_{Q,k}$ their *mean embeddings* and $p, J \geq 1$ two integers. Then we define:

$$d_{p,J}(P,Q) := \left[\frac{1}{J}\sum_{j=1}^{J} |\mu_{P,k}(\mathbf{t}_j) - \mu_{Q,k}(\mathbf{t}_j)|^p\right]^{\frac{1}{p}}$$

where $(\mathbf{t}_j)_{j=1}^{J}$ are sampled independently from any absolutely continuous Borel probability measure.

→ $d_{p,J}(\cdot,\cdot)$ is a random metric on the space of probability distributions.

Characterization of the CI

- Let $d_x, d_y, d_z \geq 1$, $\mathcal{X} := \mathbb{R}^{d_x}$, $\mathcal{Y} := \mathbb{R}^{d_y}$, and $\mathcal{Z} := \mathbb{R}^{d_z}$. Let (X,Z,Y) be a random vector on $\mathcal{X}\times\mathcal{Z}\times\mathcal{Y}$ with law P_{XZY}
- Denote $\ddot{X} := (X,Z)$, $\ddot{\mathcal{X}} := \mathcal{X}\times\mathcal{Z}$ and let us define for all mesurable $(A,B) \in \mathcal{B}(\ddot{\mathcal{X}})\times\mathcal{B}(\mathcal{Y})$:

$$P_{\ddot{X}\otimes Y|Z}(A\times B) := \mathbb{E}_Z\left[\mathbb{E}_{\ddot{X}}[\mathbf{1}_A|Z]\mathbb{E}_Y[\mathbf{1}_B|Z]\right].$$

Proposition: $d_{p,J}(P_{XZY}, P_{\ddot{X}\otimes Y|Z}) = 0$ if and only if $X \perp Y|Z$ a.s.

A First Oracle Statistic

For all $(\mathbf{t}^{(1)}, t^{(2)}) \in \ddot{\mathcal{X}}\times\mathcal{Y}$, we define the *witness function*:

$$\Delta(\mathbf{t}^{(1)}, t^{(2)}) := \mu_{P_{\ddot{X}\otimes Y|Z}, k_{\ddot{\mathcal{X}}}\cdot k_{\mathcal{Y}}}(\mathbf{t}^{(1)}, t^{(2)}) - \mu_{P_{XZY}, k_{\ddot{\mathcal{X}}}\cdot k_{\mathcal{Y}}}(\mathbf{t}^{(1)}, t^{(2)})$$

Reformulation of the witness function $\Delta(\mathbf{t}^{(1)}, t^{(2)})$:

$$\mathbb{E}\left[\left(k_{\ddot{\mathcal{X}}}(\mathbf{t}^{(1)},\ddot{X}) - \mathbb{E}_{\ddot{X}}\left[k_{\ddot{\mathcal{X}}}(\mathbf{t}^{(1)},\ddot{X})|Z\right]\right)\left(k_{\mathcal{Y}}(t^{(2)},Y) - \mathbb{E}_Y\left[k_{\mathcal{Y}}(t^{(2)},Y)|Z\right]\right)\right]$$

A First Estimate of the witness function denoted $\Delta_n(\mathbf{t}_j^{(1)}, t_j^{(2)})$:

$$\frac{1}{n}\sum_{i=1}^{n}\left(k_{\ddot{\mathcal{X}}}(\mathbf{t}^{(1)},\ddot{x}_i) - \mathbb{E}_{\ddot{X}}\left[k_{\ddot{\mathcal{X}}}(\mathbf{t}^{(1)},\ddot{X})|z_i\right]\right)\left(k_{\mathcal{Y}}(t^{(2)},y_i) - \mathbb{E}_Y\left[k_{\mathcal{Y}}(t^{(2)},Y)|z_i\right]\right)$$

Definition of Our Oracle Statistic

$$\mathrm{CI}_{n,p} := \sum_{j=1}^{J}\left|\Delta_n(\mathbf{t}_j^{(1)}, t_j^{(2)})\right|^p$$

- Under H_0, $\sqrt{n}\,\mathrm{CI}_{n,p} \to \|X\|_p^p$ where $X \sim \mathcal{N}(0_J, \Sigma)$ and we have an analytic formulation of Σ.
- Under H_1, $\lim_{n\to\infty} P(n^{p/2}\mathrm{CI}_{n,p} \geq q) = 1$ for any $q \in \mathbb{R}$.

Problems:

- The oracle statistic involves unknown conditional means
- The asymptotic distributions involved an unknown covariance

Approximation of the Oracle

We estimate these conditional means using ***Regularized Least-squares Estimators***:

$$h_{j,r}^{(2)} := \min_{h\in H_{\mathcal{Z}}^{2,j}} \frac{1}{r}\sum_{i=1}^{r}\left(h(z_i) - k_{\mathcal{Y}}(t_j^{(2)}, y_i)\right)^2 + \lambda_{j,r}^{(2)}\|h\|_{H_{\mathcal{Z}}^{2,j}}^2$$

$$h_{j,r}^{(1)} := \min_{h\in H_{\mathcal{Z}}^{1,j}} \frac{1}{r}\sum_{i=1}^{r}\left(h(z_i) - k_{\ddot{\mathcal{X}}}(\mathbf{t}_j^{(1)}, (x_i, z_i))\right)^2 + \lambda_{j,r}^{(1)}\|h\|_{H_{\mathcal{Z}}^{1,j}}^2$$

Approximate Estimate of the Witness Function

$$\widetilde{\Delta}_{n,r}(\mathbf{t}_j^{(1)}, t_j^{(2)}) := \frac{1}{n}\sum_{i=1}^{n}\left(k_{\ddot{\mathcal{X}}}(\mathbf{t}_j^{(1)},\ddot{x}_i) - h_{j,r}^{(1)}(z_i)\right)\times\left(k_{\mathcal{Y}}(t_j^{(2)},y_i) - h_{j,r}^{(2)}(z_i)\right)$$

Definition of our Approximate Statistic

$$\widetilde{\mathrm{CI}}_{n,r,p} := \sum_{j=1}^{J}\left|\widetilde{\Delta}_{n,r}(\mathbf{t}_j^{(1)}, t_j^{(2)})\right|^p$$

We show the same asymptotic behavior as the one obtained for the Oracle statistic.

Normalized Version

Denote $\tilde{u}_{i,r}(j) := (k_{\ddot{\mathcal{X}}}(\mathbf{t}_j^{(1)},\ddot{x}_i) - h_{j,r}^{(1)}(z_i))(k_{\mathcal{Y}}(t_j^{(2)},y_i) - h_{j,r}^{(2)}(z_i))$,

$$\widetilde{\mathbf{S}}_{n,r} := \frac{1}{n}\sum_{i=1}^{n}\tilde{\mathbf{u}}_{i,r} \quad\text{and}\quad \mathbf{\Sigma}_{n,r} := \frac{1}{n}\sum_{i=1}^{n}\tilde{\mathbf{u}}_{i,r}\tilde{\mathbf{u}}_{i,r}^T$$

$$\widetilde{\mathrm{NCI}}_{n,r,p} := \|\Sigma_{n,r}^{-1/2}\widetilde{\mathbf{S}}_{n,r}\|_p^p.$$

Results:

- Under H_0, $\sqrt{n}\,\widetilde{\mathrm{NCI}}_{n,r_n,p} \to \|X\|_p^p$ where $X \sim \mathcal{N}(0_J, \mathrm{Id}_J)$
- Under H_1, $\lim_{n\to\infty} P(n^{p/2}\widetilde{\mathrm{NCI}}_{n,r_n,p} \geq q) = 1$ for any $q \in \mathbb{R}$.