Gaining Power in Conditional Independence Tests

Mengqi Liu

November 17, 2023

THE UNIVERSITY OF CHICAGO

Goal

Strong CI

- $H_0 : X \perp\!\!\!\perp Y|Z$, $H_1 : X \not\perp\!\!\!\perp Y|Z$
  - Distance between $p_{x|z}\&p_{x|y,z}$ or $p_{xy|z}\&p_{x|z}p_{y|z}$
  - (Daudin 1900) $\Leftrightarrow \mathbb{E}f(X,Z)g(Y,Z) = 0$, $\forall f \in L^2_{X,Z}$, $\forall g \in L^2_{Y,Z}$, s.t. $\mathbb{E}[f(X,Z)|Z] = \mathbb{E}[g(Y,Z)|Z] = 0$.
  - conditional cross-covariance $\Sigma_{XY|Z} = \Sigma_{XY} - \Sigma_{XZ}\Sigma_{ZZ}^{-1}\Sigma_{ZY}$
  - conditional mutual information $I(X;Y|Z) = \int_Z \int_Y \int_X \log(\frac{p_{x,y|z}(x,y)}{p_{x|z}(x,z)p_{y|z}(y,z)})p_{x,y,z}(x,y,z)dxdydz$

Weak CI

- $\mathbb{E}[\mathrm{Cov}(f(X),g(Y)|Z)] = 0$
- mean independence $\mathbb{E}[Y|X,Z] = \mathbb{E}[Y|Z]$ (PCM)

Hardness in CI testing

Theorem 2 (No-free-lunch). *Given any $n \in \mathbb{N}$, $\alpha \in (0,1)$, $M \in (0,\infty]$, and any potentially randomised test ψ_n that has valid level α for the null hypothesis $\mathcal{P}_{0,M}$, we have that $\mathbb{P}_Q(\psi_n = 1) \leq \alpha$ for all $Q \in \mathcal{Q}_{0,M}$. Thus ψ_n cannot have power against any alternative.*

Corollary 3. *For all $M \in (0,\infty]$ and for any sequence $(\psi_n)_{n=1}^{\infty}$ of tests we have*

$$\sup_{Q \in \mathcal{Q}_{0,M}} \limsup_{n\to\infty} \mathbb{P}_Q(\psi_n = 1) \leq \limsup_{n\to\infty} \sup_{P \in \mathcal{P}_{0,M}} \mathbb{P}_P(\psi_n = 1).$$

- Add constraints on $\mathcal{P}$, $\mathcal{Q}$

First Case I

- $\mathcal{P}_0$ is replaced by the set of all distributions absolutely continuous with respect to Lebesgue measure
- Generalised Covariance Measure (GCM)

$$R_i = \{x_i - \hat{f}(z_i)\}\{y_i - \hat{g}(z_i)\}.$$

$$T^{(n)} = \frac{\sqrt{n} \cdot \frac{1}{n} \sum_{i=1}^{n} R_i}{\left(\frac{1}{n} \sum_{i=1}^{n} R_i^2 - \left(\frac{1}{n} \sum_{r=1}^{n} R_r\right)^2\right)^{1/2}} =: \frac{\tau_N^{(n)}}{\tau_D^{(n)}}.$$

- Estimate $\rho_P = \mathbb{E}_P \mathrm{Cov}_P(X, Y|Z)$ in $o_{\mathcal{P}}(\frac{1}{\sqrt{n}})$

$$\sup_{P \in \mathcal{P}} \sup_{t \in \mathbb{R}} \left| \mathbb{P}_P \left(\frac{\tau_N^{(n)} - \sqrt{n}\rho_P}{\tau_D^{(n)}} \leq t \right) - \Phi(t) \right| \to 0, \quad \tau_D^{(n)} - \sigma_P = o_{\mathcal{P}}(1).$$

First Case II

- Fail if
  - $\mathrm{Cov}(X, Y|Z) = 0$
  - $X \sim N(Z, \sigma^2), Y = X^2 + \epsilon$

First Case III

- Projected Covariance Measure (PCM)
  - Y is conditionally mean independent of X given Z if and only if $\mathbb{E}[\{Y - \mathbb{E}(Y|Z)\}f(X,Z)] = \mathbb{E}[\mathrm{Cov}(Y, f(X,Z)|Z)] = 0$
  - Construct test statistic

$$f(X,Z) = \frac{h(X,Z)}{v(X,Z)} = \frac{\mathbb{E}(Y \mid X,Z) - \mathbb{E}(Y \mid Z)}{\mathrm{Var}(Y \mid X,Z)}$$

$$L_i := \{Y_i - \widehat{m}(Z_i)\}\{\widehat{f}(X_i,Z_i) - \widehat{m}_{\widehat{f}}(Z_i)\},$$

$$T := \frac{\frac{1}{\sqrt{n}}\sum_{i=1}^n L_i}{\sqrt{\frac{1}{n}\sum_{i=1}^n L_i^2 - \left(\frac{1}{n}\sum_{i=1}^n L_i^2\right)}}.$$

  - Validity

$$\sup_{P\in\mathcal{P}_0}\sup_{t\in\mathbb{R}}\left|\mathbb{P}_P(T \le t) - \Phi(t)\right| \to 0.$$

First Case IV

- Power analysis

Theorem 5. *Assume that $\widehat{m}_{\widehat{f}}$ is scale equivariant in the sense that it satisfies (14) and consider the sequence of classes of distributions in Assumption 4, where*

$$\epsilon_n \cdot n^{\min\{1,\beta_1+\beta_2\}} \to \infty. \tag{15}$$

Then for any $\alpha \in (0,1)$,

$$\inf_{P \in \mathcal{P}_1(\epsilon_n)} \mathbb{P}_P(T > z_{1-\alpha}) \to 1.$$

Proposition 1. *Consider a version of the PCM setting $\widehat{v} \equiv 1$ and using OLS for each of the regressions involved for a family of distributions $\mathcal{P}$ satisfying Assumption 1 and (8). Let $\mathcal{P}_1(\kappa) := \{P \in \mathcal{P} : |\beta_P| \geq \kappa/\sqrt{n}\}$. Given any $\alpha \in (0,1)$, we have*

$$\lim_{\kappa\to\infty} \lim_{n\to\infty} \inf_{P \in \mathcal{P}_1(\kappa)} \mathbb{P}_P(T > z_{1-\alpha}) = 1.$$

Second Case I

- Considering two situations:
  - When X and Y are discrete supported on $[l_1] \times [l_2]$ for some integers l_1, l_2, and when Z has an absolutely continuous (with respect to the Lebesgue measure) distribution supported on [0,1].
  - When all three variables (X,Y,Z) have an absolutely continuous (with respect to the Lebesgue measure) distribution supported on [0,1].

Second Case II

- Define $H_0 : p \in \mathcal{H}_0$ and $H_1 : p \in S_1(\epsilon)$, where $S_1(\epsilon) = \{p \in \mathcal{H}_1 : \inf_{q \in \overline{\mathcal{H}_0}} \|p - q\| \geq \epsilon\}$.
- Minimax risk: $R_n(\mathcal{H}_0, \overline{\mathcal{H}}_0, \mathcal{H}_1, \epsilon) = \inf_\psi \{\sup_{p \in \mathcal{H}_0} \mathbb{E}_p[\psi(\mathcal{D}_n)] + \sup_{p \in S_1(\epsilon)} \mathbb{E}_p[1 - \psi(\mathcal{D}_n)]\}$
- Critical radius: $\epsilon_n(\mathcal{H}_0, \overline{\mathcal{H}}_0, \mathcal{H}_1) = \inf\{\epsilon : R_n(\mathcal{H}_0, \overline{\mathcal{H}}_0, \mathcal{H}_1, \epsilon) \leq \frac{1}{3}\}$
- result:

	X, Y		
	discrete on $[\ell_1] \times [\ell_2]$, ℓ_1, ℓ_2 fixed	discrete on $[\ell_1] \times [\ell_2]$	continuous
ε_n-Upper Bounds	$n^{-2/5}$	$\frac{(\ell_1 \ell_2)^{1/5}}{n^{2/5}}$, given $\frac{\ell_1^4}{\ell_2} \lesssim n^3$	$n^{-2s/(5s+2)}$
ε_n-Lower Bounds	$n^{-2/5}$	$\frac{(\ell_1 \ell_2)^{1/5}}{n^{2/5}}$	$n^{-2s/(5s+2)}$

Table 1: This is a summary of the minimax results obtained in the main text of our paper.

Second Case III

- s denotes the Hölder smoothness parameter of the conditional density $p_{X,Y|Z}$ under the alternative hypothesis.

Definition 2.3 (Hölder Smoothness)**.** *Let $s > 0$ be a fixed real number, and let $\lfloor s \rfloor$ denote the maximum integer strictly smaller than s. Denote by $\mathcal{H}^{2,s}(L)$, the class of functions $f : [0,1]^2 \mapsto \mathbb{R}$, which posses all partial derivatives up to order $\lfloor s \rfloor$ and for all $x, y, x', y' \in [0,1]$ we have*

$$\sup_{k \le \lfloor s \rfloor} \left| \frac{\partial^k}{\partial x^k} \frac{\partial^{\lfloor s \rfloor - k}}{\partial y^{\lfloor s \rfloor - k}} f(x,y) - \frac{\partial^k}{\partial x^k} \frac{\partial^{\lfloor s \rfloor - k}}{\partial y^{\lfloor s \rfloor - k}} f(x',y') \right| \le L((x-x')^2 + (y-y')^2))^{\frac{s-\lfloor s \rfloor}{2}}, \tag{2.4}$$

and in addition

$$\sup_{k \le \lfloor s \rfloor} \left| \frac{\partial^k}{\partial x^k} \frac{\partial^{\lfloor s \rfloor - k}}{\partial y^{\lfloor s \rfloor - k}} f(x,y) \right| \le L.$$

Second Case IV

- Constructing test statistic T (continuous X, Y, Z)
  - Draw $N \sim Poi(\frac{n}{2})$ and take arbitrary N out of the n observations in the case when $Nleqn$, and accept the null hypothesis if $N > n$.
  - $[0,1] = \cup_{i\in[d]} C_i = \cup_{i\in[d']} C_i^{'}$, $d' = \lceil d^{1/s} \rceil$, C_i's discretize Z, $C_i^{'}$'s discretize X and $Y \longrightarrow \mathcal{D}'_N$
  - $\phi_{ij}(xy) = \mathbb{1}(X'_i = x, Y'_j = y) - \mathbb{1}(X'_i = x)\mathbb{1}(Y'_j = y)$
  - $h_{ijkl} = \frac{1}{4!} \sum_{\pi\in[4!]} \sum_{x\in[d'],y\in[d']} \phi_{\pi_1\pi_2}(xy)\phi_{\pi_3\pi_4}(xy)$
  - $U(\mathcal{D}) = \frac{1}{\binom{\sigma}{4}} \sum_{i<j<k<l:(i,j,k,l)\in[\sigma]} h_{ijkl}$
  - $T = \sum_{m\in[d]} \mathbb{1}(\sigma_m \geq 4)\sigma_m U(\mathcal{D}'_m)$

Second Case V

- Power analysis with $\psi_\tau(\mathcal{D}_N) = \mathbb{1}(T \geq \tau)$

Theorem 5.6 (Continuous X, Y, Z Upper Bound). *Set $d = \lceil n^{2s/(5s+2)} \rceil$ and set the threshold $\tau = \sqrt{\zeta d}$ for a sufficiently large ζ (depending on L). Let $\mathcal{H}_0(s) = \mathcal{P}_{0,[0,1]^3,\mathrm{TV}}(L) \cup \mathcal{P}_{0,[0,1]^3,\chi^2}(L)$ when $s \geq 1$ and $\mathcal{H}_0(s) = \mathcal{P}_{0,[0,1]^3,\chi^2}(L)$ when $s < 1$. Then, for a sufficiently large absolute constant c (depending on ζ, L), when $\varepsilon \geq cn^{-2s/(5s+2)}$, we have that*

$$\sup_{p \in \mathcal{H}_0(s)} \mathbb{E}_p[\psi_\tau(\mathcal{D}'_k)] \leq \frac{1}{10},$$

$$\sup_{p \in \{p \in \mathcal{Q}_{0,[0,1]^3,\mathrm{TV}}(L,s) : \inf_{q \in \mathcal{P}_{0,[0,1]^3}} \|p-q\|_1 \geq \varepsilon\}} \mathbb{E}_p[1 - \psi_\tau(\mathcal{D}'_k)] \leq \frac{1}{10} + \exp(-n/8).$$

Second Case VI

Local Permutation Test

Algorithm 1 Local permutation procedure

Input: data $\{(X_i, Y_i, Z_i)\}_{i=1}^n$, a partition of $\mathcal{Z}$: $\{B_1, \ldots, B_M\}$, a test statistic T_{CI}, a nominal level α

1. For each $\boldsymbol{\pi} \in \boldsymbol{\Pi}$, compute $T_{\mathrm{CI}}^{\boldsymbol{\pi}}$ as in (2) and denote the resulting statistics by $T_{\mathrm{CI}}^{\boldsymbol{\pi}_1}, \ldots, T_{\mathrm{CI}}^{\boldsymbol{\pi}_K}$.
2. By comparing the statistic T_{CI} in (1) with the permuted ones, calculate the p-value as

$$p_{\mathrm{perm}} = \frac{1}{K} \sum_{\boldsymbol{\pi}_i \in \boldsymbol{\Pi}} \mathbb{1}\{T_{\mathrm{CI}}^{\boldsymbol{\pi}_i} \geq T_{\mathrm{CI}}\}. \tag{3}$$

3. Given the nominal level $\alpha \in (0, 1)$, define the test function $\phi_{\mathrm{perm},n} = \mathbb{1}(p_{\mathrm{perm}} \leq \alpha)$ and reject the null when $\phi_{\mathrm{perm},n} = 1$.

Second Case VII

- Review of hardness (weaker condition)

Theorem 1 (Hardness of CI testing). *For an arbitrary integer $J \geq n(n-1)$, let us define $\rho_{J,P} := \mathbb{P}\{Z_1, \ldots, Z_J$ are distinct$\}$, where $Z_1, \ldots, Z_J$ are i.i.d. samples from the marginal distribution of Z. Suppose that a test ϕ satisfies $\sup_{P_{X,Y,Z} \in \mathcal{P}_{0,\mathrm{disc}}} \mathbb{E}_{P^n_{X,Y,Z}}[\phi] \leq \alpha$ for $\alpha \in (0,1)$. Then for any $P_{X,Y,Z} \in \mathcal{P}_1$, the power of ϕ is bounded above by*

$$\mathbb{E}_{P^n_{X,Y,Z}}[\phi] \leq \alpha \times \rho_{J,P} + (1 - \rho_{J,P}) + \frac{n(n-1)}{J}. \tag{4}$$

- $M = \lceil n^{2s/(5s+2)} \rceil$, $\epsilon \geq cn^{-2s/(5s+2)}$
- type-II error guarantee

$$\sup_{P_{X,Y,Z} \in \mathcal{P}_{1,[0,1]^3,\mathrm{TV}}(L) : \inf_{Q \in \mathcal{P}_{0,[0,1]^3}} \mathcal{D}_{\mathrm{TV}}(P_{X,Y,Z}, Q) \geq \varepsilon} \mathbb{E}_{P^N_{X,Y,Z}, N}[1 - \phi_{\mathrm{NBW},2}] \leq \frac{1}{100} + e^{-n/8}. \tag{21}$$

Reference

Rajen D. Shah and Jonas Peters.
The hardness of conditional independence testing and the generalised covariance measure.
The Annals of Statistics, 48(3), June 2020.

Anton Rask Lundborg, Ilmun Kim, Rajen D. Shah, and Richard J. Samworth.
The projected covariance measure for assumption-lean variable significance testing, 2023.

Matey Neykov, Sivaraman Balakrishnan, and Larry Wasserman.
Minimax optimal conditional independence testing, 2021.

Ilmun Kim, Matey Neykov, Sivaraman Balakrishnan, and Larry Wasserman.
Local permutation tests for conditional independence, 2022.

Chun Li and Xiaodan Fan.
On nonparametric conditional independence tests for continuous variables.
WIREs Computational Statistics, 12(3):e1489, 2020.

Thank You!