The Projected Covariance Measure for model–free variable significance testing

Anton Rask Lundborg

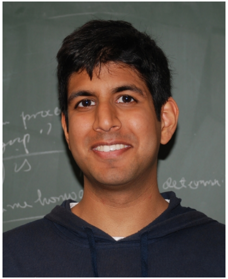
ETH–UCPH–TUM Workshop

October 2022

Outline

- Formalising variable significance
- Regression-based variable significance tests
- The Projected Covariance Measure (PCM)
- Numerical results

Collaborators

Ilmun Kim — Yonsei University

Rajen Shah, Richard Samworth — University of Cambridge

Introduction

Understanding the relation between a response and associated predictors, and selecting those predictors that are important, is a common problem faced by statisticians and data analysts.

Introduction

Understanding the relation between a response and associated predictors, and selecting those predictors that are important, is a common problem faced by statisticians and data analysts.

When $(Y, X, Z) \in \mathbb{R} \times \mathbb{R}^{d_X} \times \mathbb{R}^{d_Z}$ a simple but popular way of addressing this is to fit a linear model

$$Y = \beta^\top X + \gamma^\top Z + \varepsilon, \quad \mathbb{E}(\varepsilon \,|\, X, Z) = 0,$$

and perform an F-test for the significance of X.

Introduction

Understanding the relation between a response and associated predictors, and selecting those predictors that are important, is a common problem faced by statisticians and data analysts.

When $(Y, X, Z) \in \mathbb{R} \times \mathbb{R}^{d_X} \times \mathbb{R}^{d_Z}$ a simple but popular way of addressing this is to fit a linear model

$$Y = \beta^\top X + \gamma^\top Z + \varepsilon, \quad \mathbb{E}(\varepsilon \,|\, X, Z) = 0,$$

and perform an F-test for the significance of X.

When the linear model is misspecified, we might either wrongly declare X to be important or unimportant, and similar issues arise from other tests based on parametric models.

What does it mean for a variable to be significant?

These issues combined with the increasing use and effectiveness of nonparametric methods lead us to require a model-free hypothesis.

What does it mean for a variable to be significant?

These issues combined with the increasing use and effectiveness of nonparametric methods lead us to require a model-free hypothesis.

We consider conditional mean independence; real-valued Y is conditionally mean independent of X given Z if

$$\mathbb{E}(Y \mid X, Z) = \mathbb{E}(Y \mid Z).$$

What does it mean for a variable to be significant?

These issues combined with the increasing use and effectiveness of nonparametric methods lead us to require a model-free hypothesis.

We consider conditional mean independence; real-valued Y is conditionally mean independent of X given Z if

$$\mathbb{E}(Y \mid X, Z) = \mathbb{E}(Y \mid Z).$$

The alternative, conditional mean dependence, may be characterised by the property that X improves the prediction of Y in a mean-squared error sense, given knowledge of Z.

Contrasting with conditional independence

A more common model-free hypothesis is that of conditional independence; we say that Y and X are conditionally independent given Z and write $Y \perp\!\!\!\perp X \mid Z$ if

$$\mathbb{E}(f(Y) \mid X, Z) = \mathbb{E}(f(Y) \mid Z)$$

for all suitable real-valued functions f.

Contrasting with conditional independence

A more common model-free hypothesis is that of conditional independence; we say that Y and X are conditionally independent given Z and write $Y \perp\!\!\!\perp X \,|\, Z$ if

$$\mathbb{E}(f(Y) \,|\, X, Z) = \mathbb{E}(f(Y) \,|\, Z)$$

for all suitable real-valued functions f.

Any test of conditional mean independence may also be used as a test of conditional independence, although it will not be powerful against all alternatives.

Contrasting with conditional independence

A more common model-free hypothesis is that of conditional independence; we say that Y and X are conditionally independent given Z and write $Y \perp\!\!\!\perp X \,|\, Z$ if

$$\mathbb{E}(f(Y) \,|\, X, Z) = \mathbb{E}(f(Y) \,|\, Z)$$

for all suitable real-valued functions f.

Any test of conditional mean independence may also be used as a test of conditional independence, although it will not be powerful against all alternatives.

This also means we are faced with the same statistical limitations as when testing conditional independence.

The hardness of testing conditional (mean) independence

Testing for conditional independence is a difficult problem without further assumptions.

The hardness of testing conditional (mean) independence

Testing for conditional independence is a difficult problem without further assumptions.

Suppose $(Y, X, Z) \in \mathbb{R}^3$ and the joint distribution has a density with respect to Lebesgue measure. Then, any test that rejects with probability α under the null, has power at most α against any alternative distribution [Shah and Peters, 2020].

The hardness of testing conditional (mean) independence

Testing for conditional independence is a difficult problem without further assumptions.

Suppose $(Y, X, Z) \in \mathbb{R}^3$ and the joint distribution has a density with respect to Lebesgue measure. Then, any test that rejects with probability α under the null, has power at most α against any alternative distribution [Shah and Peters, 2020].

As a consequence of this result, we know that domain knowledge is required to select a conditional independence test tailored to the problem at hand.

The Generalised Covariance Measure (GCM)

The Generalised Covariance Measure (GCM) is a conditional (mean) independence test relying (primarily) on the ability of user-chosen regression methods for estimating conditional expectations [Shah and Peters, 2020].

The Generalised Covariance Measure (GCM)

The Generalised Covariance Measure (GCM) is a conditional (mean) independence test relying (primarily) on the ability of user-chosen regression methods for estimating conditional expectations [Shah and Peters, 2020].

For $X \in \mathbb{R}$, set

$$L_i := \{Y_i - \hat{m}_{Y|Z}(Z_i)\}\{X_i - \hat{m}_{X|Z}(Z_i)\}$$

$$\mathrm{GCM}_{Y,X|Z} := \sqrt{n}\frac{\frac{1}{n}\sum_{i=1}^{n} L_i}{\left\{\frac{1}{n}\sum_{i=1}^{n}(L_i - \bar{L})^2\right\}^{1/2}}.$$

The Generalised Covariance Measure (GCM)

The Generalised Covariance Measure (GCM) is a conditional (mean) independence test relying (primarily) on the ability of user-chosen regression methods for estimating conditional expectations [Shah and Peters, 2020].

For $X \in \mathbb{R}$, set

$$L_i := \{Y_i - \hat{m}_{Y|Z}(Z_i)\}\{X_i - \hat{m}_{X|Z}(Z_i)\}$$

$$\mathrm{GCM}_{Y,X|Z} := \sqrt{n}\frac{\frac{1}{n}\sum_{i=1}^n L_i}{\left\{\frac{1}{n}\sum_{i=1}^n (L_i - \bar{L})^2\right\}^{1/2}}.$$

Under conditions $\mathrm{GCM}_{Y,X|Z} \xrightarrow{d} \mathcal{N}(0,1)$ under the null.

The Generalised Covariance Measure (GCM)

The Generalised Covariance Measure (GCM) is a conditional (mean) independence test relying (primarily) on the ability of user-chosen regression methods for estimating conditional expectations [Shah and Peters, 2020].

For $X \in \mathbb{R}$, set

$$L_i := \{Y_i - \hat{m}_{Y|Z}(Z_i)\}\{X_i - \hat{m}_{X|Z}(Z_i)\}$$

$$\mathrm{GCM}_{Y,X|Z} := \sqrt{n}\frac{\frac{1}{n}\sum_{i=1}^n L_i}{\left\{\frac{1}{n}\sum_{i=1}^n (L_i - \bar{L})^2\right\}^{1/2}}.$$

Under conditions $\mathrm{GCM}_{Y,X|Z} \xrightarrow{d} \mathcal{N}(0,1)$ under the null.

The primary requirement is that

$$\frac{1}{n}\sum_{i=1}^n \{m_{Y|Z}(Z_i) - \hat{m}_{Y|Z}(Z_i)\}^2 \cdot \frac{1}{n}\sum_{i=1}^n \{m_{X|Z}(Z_i) - \hat{m}_{X|Z}(Z_i)\}^2 = o_P(n^{-1}).$$

Power of the GCM

As the GCM is a normalised version of $\mathbb{E}\mathrm{Cov}(Y, X \mid Z)$, we only have power when $\mathbb{E}\mathrm{Cov}(Y, X \mid Z) \neq 0$, which is not always the case when $\mathbb{E}(Y \mid X, Z) \neq \mathbb{E}(Y \mid Z)$.

Power of the GCM

As the GCM is a normalised version of $\mathbb{E}\mathrm{Cov}(Y, X \mid Z)$, we only have power when $\mathbb{E}\mathrm{Cov}(Y, X \mid Z) \neq 0$, which is not always the case when $\mathbb{E}(Y \mid X, Z) \neq \mathbb{E}(Y \mid Z)$.

Consider $(X, Z, \varepsilon) \sim \mathcal{N}_3(0, I)$ and $Y = X^2 + \varepsilon$. Here, $\mathrm{Cov}(Y, X \mid Z) = 0$ so $\mathbb{E}\{\mathrm{Cov}(Y, X \mid Z)\} = 0$ hence the GCM is powerless.

Power of the GCM

As the GCM is a normalised version of $\mathbb{E}\mathrm{Cov}(Y, X \mid Z)$, we only have power when $\mathbb{E}\mathrm{Cov}(Y, X \mid Z) \neq 0$, which is not always the case when $\mathbb{E}(Y \mid X, Z) \neq \mathbb{E}(Y \mid Z)$.

Consider $(X, Z, \varepsilon) \sim \mathcal{N}_3(0, I)$ and $Y = X^2 + \varepsilon$. Here, $\mathrm{Cov}(Y, X \mid Z) = 0$ so $\mathbb{E}\{\mathrm{Cov}(Y, X \mid Z)\} = 0$ hence the GCM is powerless.

Scheidegger et al. [2021] introduce a carefully weighted version of the GCM that can have power when $\mathrm{Cov}(Y, X \mid Z) \neq 0$, but this remains powerless in the above example.

Power of the GCM

As the GCM is a normalised version of $\mathbb{E}\mathrm{Cov}(Y, X \mid Z)$, we only have power when $\mathbb{E}\mathrm{Cov}(Y, X \mid Z) \neq 0$, which is not always the case when $\mathbb{E}(Y \mid X, Z) \neq \mathbb{E}(Y \mid Z)$.

Consider $(X, Z, \varepsilon) \sim \mathcal{N}_3(0, I)$ and $Y = X^2 + \varepsilon$. Here, $\mathrm{Cov}(Y, X \mid Z) = 0$ so $\mathbb{E}\{\mathrm{Cov}(Y, X \mid Z)\} = 0$ hence the GCM is powerless.

Scheidegger et al. [2021] introduce a carefully weighted version of the GCM that can have power when $\mathrm{Cov}(Y, X \mid Z) \neq 0$, but this remains powerless in the above example.

We would like a test that:

- relies primarily on user-chosen machine learning methods performing sufficiently well (i.e. restricts the null in this fashion);
- has power against a more diverse set of alternatives.

Estimating the mean squared difference in regression functions

Williamson et al. [2021a] propose to estimate

$$\tau := \mathbb{E}[\{\mathbb{E}(Y \mid X, Z) - \mathbb{E}(Y \mid Z)\}^2] = \mathbb{E}[\{Y - \mathbb{E}(Y \mid Z)\}^2] - \mathbb{E}[\{Y - \mathbb{E}(Y \mid X, Z)\}^2]$$

via

$$\hat{\tau} := \frac{1}{n}\sum_{i=1}^{n}\{Y_i - \hat{m}_{Y|Z}(Z_i)\}^2 - \frac{1}{n}\sum_{i=1}^{n}\{Y_i - \hat{m}_{Y|X,Z}(X_i, Z_i)\}^2.$$

Estimating the mean squared difference in regression functions

Williamson et al. [2021a] propose to estimate

$$\tau := \mathbb{E}[\{\mathbb{E}(Y \mid X, Z) - \mathbb{E}(Y \mid Z)\}^2] = \mathbb{E}[\{Y - \mathbb{E}(Y \mid Z)\}^2] - \mathbb{E}[\{Y - \mathbb{E}(Y \mid X, Z)\}^2]$$

via

$$\hat{\tau} := \frac{1}{n}\sum_{i=1}^{n}\{Y_i - \hat{m}_{Y|Z}(Z_i)\}^2 - \frac{1}{n}\sum_{i=1}^{n}\{Y_i - \hat{m}_{Y|X,Z}(X_i, Z_i)\}^2.$$

$\hat{\tau}$ is asymptotically Gaussian centered on τ and achieves the semiparametric efficient variance bound provided $\tau > 0$.

Estimating the mean squared difference in regression functions

Williamson et al. [2021a] propose to estimate

$$\tau := \mathbb{E}[\{\mathbb{E}(Y \mid X, Z) - \mathbb{E}(Y \mid Z)\}^2] = \mathbb{E}[\{Y - \mathbb{E}(Y \mid Z)\}^2] - \mathbb{E}[\{Y - \mathbb{E}(Y \mid X, Z)\}^2]$$

via

$$\hat{\tau} := \frac{1}{n}\sum_{i=1}^{n}\{Y_i - \hat{m}_{Y|Z}(Z_i)\}^2 - \frac{1}{n}\sum_{i=1}^{n}\{Y_i - \hat{m}_{Y|X,Z}(X_i, Z_i)\}^2.$$

$\hat{\tau}$ is asymptotically Gaussian centered on τ and achieves the semiparametric efficient variance bound provided $\tau > 0$.

However, the functional τ is not pathwise differentiable at distributions where $\tau = 0$, so classical semiparametric theory is not applicable to the problem of testing $\tau = 0$. Consequently, $\sqrt{n}\hat{\tau}$ becomes degenerate under the null.

Estimating the mean squared difference in regression functions

Williamson et al. [2021a] propose to estimate

$$\tau := \mathbb{E}[\{\mathbb{E}(Y \mid X, Z) - \mathbb{E}(Y \mid Z)\}^2] = \mathbb{E}[\{Y - \mathbb{E}(Y \mid Z)\}^2] - \mathbb{E}[\{Y - \mathbb{E}(Y \mid X, Z)\}^2]$$

via

$$\hat{\tau} := \frac{1}{n}\sum_{i=1}^{n}\{Y_i - \hat{m}_{Y|Z}(Z_i)\}^2 - \frac{1}{n}\sum_{i=1}^{n}\{Y_i - \hat{m}_{Y|X,Z}(X_i, Z_i)\}^2.$$

$\hat{\tau}$ is asymptotically Gaussian centered on τ and achieves the semiparametric efficient variance bound provided $\tau > 0$.

However, the functional τ is not pathwise differentiable at distributions where $\tau = 0$, so classical semiparametric theory is not applicable to the problem of testing $\tau = 0$. Consequently, $\sqrt{n}\hat{\tau}$ becomes degenerate under the null.

Williamson et al. [2021b] propose a variant involving sample-splitting, but this approach sacrifices power.

An alternative approach

Our approach is based on the following characterisation of conditional mean independence:

$$\mathbb{E}(\{Y - \mathbb{E}(Y|Z)\}f(X, Z)) = 0 \quad \text{for all suitable } f.$$

An alternative approach

Our approach is based on the following characterisation of conditional mean independence:

$$\mathbb{E}(\{Y - \mathbb{E}(Y|Z)\}f(X,Z)) = 0 \quad \text{for all suitable } f.$$

Consider the following oracular test statistic. Set $L_i^* := \{Y_i - \mathbb{E}(Y_i \mid Z_i)\}f(X_i, Z_i)$ and

$$T^* := \frac{\frac{1}{\sqrt{n}} \sum_{i=1}^n L_i^*}{\sqrt{\frac{1}{n} \sum_{i=1}^n (\tilde{L}_i^*)^2}},$$

where $\tilde{L}_i^* := \{Y_i - \mathbb{E}(Y_i \mid X_i, Z_i)\}f(X_i, Z_i)$.

An alternative approach

Our approach is based on the following characterisation of conditional mean independence:

$$\mathbb{E}(\{Y - \mathbb{E}(Y|Z)\}f(X,Z)) = 0 \quad \text{for all suitable } f.$$

Consider the following oracular test statistic. Set $L_i^* := \{Y_i - \mathbb{E}(Y_i \mid Z_i)\}f(X_i, Z_i)$ and

$$T^* := \frac{\frac{1}{\sqrt{n}} \sum_{i=1}^n L_i^*}{\sqrt{\frac{1}{n} \sum_{i=1}^n (\tilde{L}_i^*)^2}},$$

where $\tilde{L}_i^* := \{Y_i - \mathbb{E}(Y_i \mid X_i, Z_i)\}f(X_i, Z_i)$.

$\mathbb{E}L_i^* / \sqrt{\mathrm{Var}(\tilde{L}_i^*)} \approx \mathbb{E}T^*$ is maximised under the alternative via

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

The Projected Covariance Measure (PCM)

Using these ideas, we propose the Projected Covariance Measure (PCM):

1. Split the sample randomly into $\mathcal{D}_1$ and $\mathcal{D}_2$.

The Projected Covariance Measure (PCM)

Using these ideas, we propose the Projected Covariance Measure (PCM):

1. Split the sample randomly into $\mathcal{D}_1$ and $\mathcal{D}_2$.
2. Produce an estimate of v and of h, $\widehat{h}$ and $\widehat{v}$, using $\mathcal{D}_2$ and set $\widehat{f}(x, z) := \widehat{h}(x, z)/\widehat{v}(x, z)$.

The Projected Covariance Measure (PCM)

Using these ideas, we propose the Projected Covariance Measure (PCM):

1. Split the sample randomly into $\mathcal{D}_1$ and $\mathcal{D}_2$.
2. Produce an estimate of v and of h, $\widehat{h}$ and $\widehat{v}$, using $\mathcal{D}_2$ and set $\widehat{f}(x, z) := \widehat{h}(x, z)/\widehat{v}(x, z)$.
3. Set $T := \mathrm{GCM}_{Y, \widehat{f}(X,Z)|Z}$, computed on $\mathcal{D}_1$, and reject when $T > z_{1-\alpha}$.

The Projected Covariance Measure (PCM)

Using these ideas, we propose the Projected Covariance Measure (PCM):

1. Split the sample randomly into $\mathcal{D}_1$ and $\mathcal{D}_2$.
2. Produce an estimate of v and of h, $\widehat{h}$ and $\widehat{v}$, using $\mathcal{D}_2$ and set $\widehat{f}(x, z) := \widehat{h}(x, z)/\widehat{v}(x, z)$.
3. Set $T := \mathrm{GCM}_{Y,\widehat{f}(X,Z)|Z}$, computed on $\mathcal{D}_1$, and reject when $T > z_{1-\alpha}$.

$h(X, Z) = 0$ under the null, so both the numerator and denominator of the test converges to 0!

The Projected Covariance Measure (PCM)

Using these ideas, we propose the Projected Covariance Measure (PCM):

1. Split the sample randomly into $\mathcal{D}_1$ and $\mathcal{D}_2$.
2. Produce an estimate of v and of h, $\widehat{h}$ and $\widehat{v}$, using $\mathcal{D}_2$ and set $\widehat{f}(x, z) := \widehat{h}(x, z)/\widehat{v}(x, z)$.
3. Set $T := \mathrm{GCM}_{Y, \widehat{f}(X,Z)|Z}$, computed on $\mathcal{D}_1$, and reject when $T > z_{1-\alpha}$.

$h(X, Z) = 0$ under the null, so both the numerator and denominator of the test converges to 0!

Despite this, the primary condition for Type I error control is

$$\frac{1}{n}\sum_{i=1}^{n}\{m_{Y|Z}(Z_i) - \hat{m}_{Y|Z}(Z_i)\}^2 \cdot \frac{1}{n\sigma^2}\sum_{i=1}^{n}\{m_{\widehat{f}|Z}(Z_i) - \hat{m}_{\widehat{f}|Z}(Z_i)\}^2 = o_P(n^{-1}),$$

where $\sigma := \mathrm{Var}(\widehat{f}(X, Z) - m_{\widehat{f}|Z}(Z) \,|\, \widehat{f})$.

Power of the PCM

Under suitable s-Hölder smoothness conditions on certain nuisance functions, no test can have power against all alternatives with

$$\tau \lesssim n^{-4s/(4s+d_X+d_Z)}.$$

Power of the PCM

Under suitable s-Hölder smoothness conditions on certain nuisance functions, no test can have power against all alternatives with

$$\tau \lesssim n^{-4s/(4s+d_X+d_Z)}.$$

By employing additional sample-splitting as in Newey and Robins [2018], setting $\widehat{v} \equiv 1$ for simplicity and using regression splines for each of our regression, we obtain (under conditions) that

- $T \xrightarrow{d} \mathcal{N}(0,1)$ under the null;

Power of the PCM

Under suitable s-Hölder smoothness conditions on certain nuisance functions, no test can have power against all alternatives with

$$\tau \lesssim n^{-4s/(4s+d_X+d_Z)}.$$

By employing additional sample-splitting as in Newey and Robins [2018], setting $\widehat{v} \equiv 1$ for simplicity and using regression splines for each of our regression, we obtain (under conditions) that

- $T \xrightarrow{d} \mathcal{N}(0, 1)$ under the null;
- The PCM has uniform power against alternatives with

$$\tau \gtrsim n^{-4s/(4s+d_X+d_Z)}.$$

The PCM is thus minimax optimal in this setting.

Numerical results

Let $Z \in \mathbb{R}^7$, $\varepsilon, \xi \in \mathbb{R}$ be independent, with $(Z, \varepsilon, \xi) \sim \mathcal{N}_9(0, I)$ and consider the null setting where

$$X = \sin(2\pi Z_1)(1 + Z_2) + \xi, \quad Y = \sin(2\pi Z_1)(1 + Z_2) + \varepsilon.$$

Numerical results

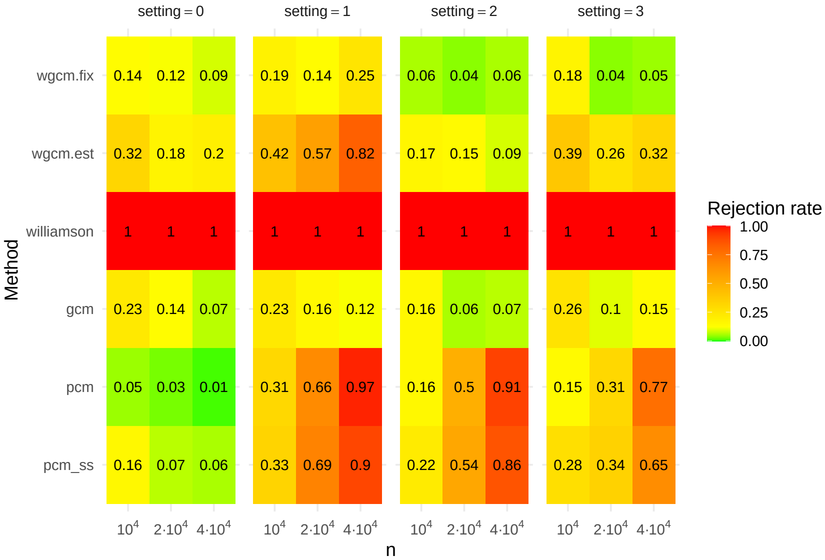
Let $Z \in \mathbb{R}^7$, $\varepsilon, \xi \in \mathbb{R}$ be independent, with $(Z, \varepsilon, \xi) \sim \mathcal{N}_9(0, I)$ and consider the null setting where

$$X = \sin(2\pi Z_1)(1 + Z_2) + \xi, \quad Y = \sin(2\pi Z_1)(1 + Z_2) + \varepsilon.$$

Consider also alternative settings, that are modifications of the null, where

1. $\mathbb{E}\mathrm{Cov}(Y, X \mid Z) = 0$ but $\mathrm{Cov}(Y, X \mid Z) \neq 0$ (additive effect);
2. $\mathrm{Cov}(Y, X \mid Z) = 0$ but $\mathbb{E}(Y \mid X, Z) \neq \mathbb{E}(Y \mid Z)$;
3. $\mathbb{E}\mathrm{Cov}(Y, X \mid Z) = 0$ but $\mathrm{Cov}(Y, X \mid Z) \neq 0$ (interaction effect);

PCM simulations using ranger



Conclusion

- The PCM works by first finding a 'projection' that is expected to expose signal in the residuals from regressing on Z on one part of the data. On the second part, we compute a generalised covariance measure statistic using this projection.

Conclusion

- The PCM works by first finding a ‘projection’ that is expected to expose signal in the residuals from regressing on Z on one part of the data. On the second part, we compute a generalised covariance measure statistic using this projection.
- Uniform type I error control is guaranteed is settings ranging from high-dimensional to nonparametric.

Conclusion

- The PCM works by first finding a ‘projection’ that is expected to expose signal in the residuals from regressing on Z on one part of the data. On the second part, we compute a generalised covariance measure statistic using this projection.
- Uniform type I error control is guaranteed is settings ranging from high-dimensional to nonparametric.
- Delivers minimax optimal power in nonparametric settings.

Conclusion

- The PCM works by first finding a ‘projection’ that is expected to expose signal in the residuals from regressing on Z on one part of the data. On the second part, we compute a generalised covariance measure statistic using this projection.
- Uniform type I error control is guaranteed is settings ranging from high-dimensional to nonparametric.
- Delivers minimax optimal power in nonparametric settings.
- Paper and R–package coming soon!

Conclusion

- The PCM works by first finding a 'projection' that is expected to expose signal in the residuals from regressing on Z on one part of the data. On the second part, we compute a generalised covariance measure statistic using this projection.
- Uniform type I error control is guaranteed is settings ranging from high-dimensional to nonparametric.
- Delivers minimax optimal power in nonparametric settings.
- Paper and R–package coming soon!

Thank you for listening.

References

Whitney K. Newey and James R. Robins. Cross-fitting and fast remainder rates for semiparametric estimation. arXiv, 2018.

Cyrill Scheidegger, Julia Hörrmann, and Peter Bühlmann. The weighted generalised covariance measure. arXiv, 2021.

Rajen D. Shah and Jonas Peters. The hardness of conditional independence testing and the generalised covariance measure. The Annals of Statistics, 48(3):1514 – 1538, 2020. doi: 10.1214/19-AOS1857. URL https://doi.org/10.1214/19-AOS1857.

Brian D. Williamson, Peter B. Gilbert, Marco Carone, and Noah Simon. Nonparametric variable importance assessment using machine learning techniques. Biometrics, 2021a. doi: https://doi.org/10.1111/biom.13392. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/biom.13392.

Brian D. Williamson, Peter B. Gilbert, Noah R. Simon, and Marco Carone. A general framework for inference on algorithm-agnostic variable importance. Journal of the American Statistical Association, 2021b. doi: 10.1080/01621459.2021.2003200. URL https://doi.org/10.1080/01621459.2021.2003200.