# Conditional Independence Testing with Applications to Causal Inference

Rajen D. Shah (Statistical Laboratory, University of Cambridge)

StatScale Workshop
20 April 2023

Jonas Peters

Ilmun Kim

Anton Lundborg

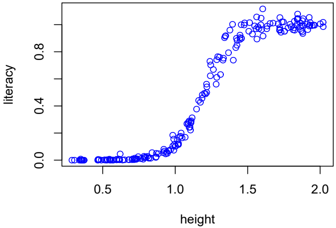
Richard Samworth

# Conditional independence

We often want to understand which variables are ‘related’ in a dataset.

Statistical formalisms of *(un)correlatedness* or *(in)dependence* can be less useful.

| Var 1 | Var 2 | ... | Var $p$ |
|---|---|---|---|
| 0.53 | 1.95 | ... | 5.21 |
| 1.32 | 3.47 | ... | 8.31 |



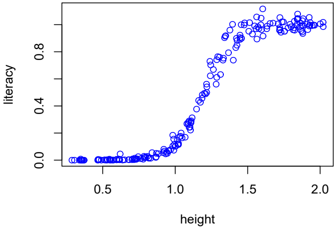

E.g., ‘Height’ and ‘literacy’ would not be independent...

# Conditional independence

We often want to understand which variables are 'related' in a dataset.

Statistical formalisms of *(un)correlatedness* or *(in)dependence* can be less useful.

| Var 1 | Var 2 | ... | Var $p$ |
|---|---|---|---|
| 0.53 | 1.95 | ... | 5.21 |
| 1.32 | 3.47 | ... | 8.31 |





E.g., 'Height' and 'literacy' would not be independent...

...but they would expected to be *conditionally independent* given 'age'.

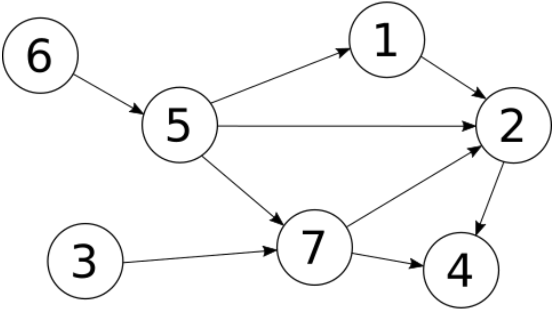
$X$ being conditionally independent of $Y$, given $Z$, written $X \perp\!\!\!\perp Y \mid Z$, encapsulates the notion that "knowing $Z$ renders $X$ irrelevant for predicting $Y$." (Lauritzen, 1991)

# Outline

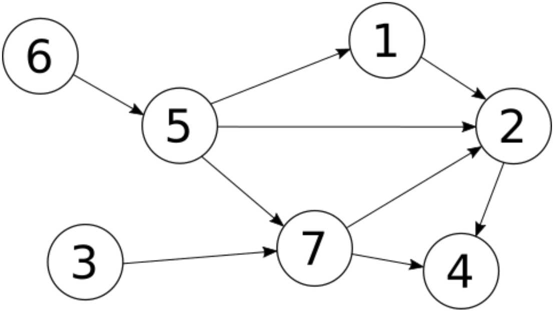
- Applications of conditional independence
- Testing conditional independence
- Generalised Covariance Measures
- Numerical results

# Applications of conditional independence

# Structural Causal Models

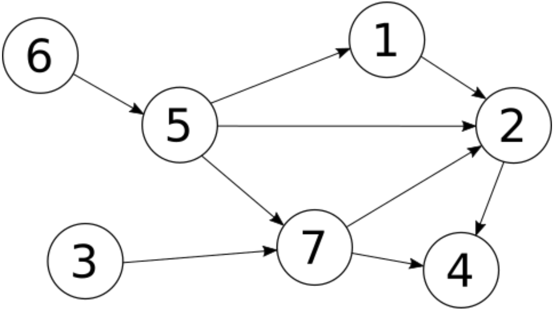


# Structural Causal Models




$$X_5 = f_5(X_6, \varepsilon_5)$$
$$X_7 = f_7(X_3, X_5, \varepsilon_7)$$
$$X_1 = f_1(X_5, \varepsilon_1)$$
$$X_2 = f_2(X_1, X_5, X_7, \varepsilon_2)$$
$$X_4 = f_4(X_2, X_7, \varepsilon_4).$$

# Structural Causal Models

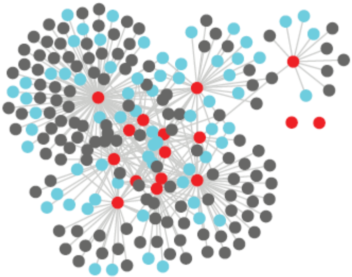



$$
\begin{aligned}
X_5 &= f_5(X_6, \varepsilon_5) \\
X_7 &= f_7(X_3, X_5, \varepsilon_7) \\
X_1 &= f_1(X_5, \varepsilon_1) \\
X_2 &= f_2(X_1, X_5, X_7, \varepsilon_2) \\
X_4 &= f_4(X_2, X_7, \varepsilon_4).
\end{aligned}
$$

If $X_j$ and $X_k$ are not connected, they must be conditionally independent given some subset of the variables.

# Conditional independence graphs

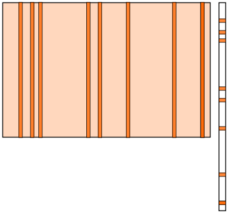
Conditional independence graphs include an edge $X_j - X_k$ when $X_j \not\!\perp\!\!\!\perp X_k \mid X_{-jk}$.



If $X_j - X_k$ then either $X_j \to X_k$, $X_k \to X_j$ or $X_j \to X_l \leftarrow X_k$.

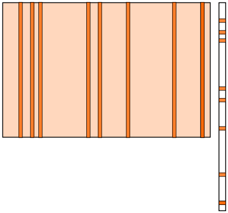
$X_j$ and $X_k$ are 'at most one step away from being causally related".

# Variable importance



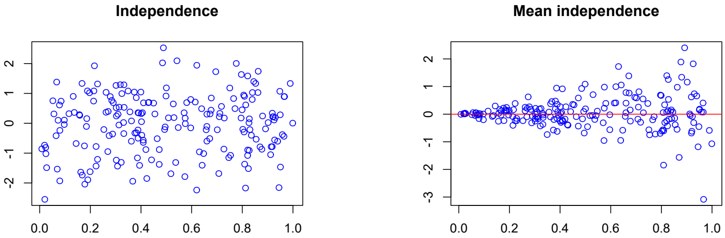
$Y \not\!\perp\!\!\!\perp X_j \,|\, X_{-j}$ gives a model-free way of formalising '$X_j$ is important for learning about $Y$, given $X_{-j}$'.

# Variable importance



$Y \not\!\perp\!\!\!\perp X_j \,|\, X_{-j}$ gives a model-free way of formalising '$X_j$ is important for learning about $Y$, given $X_{-j}$'.

Related to conditional *mean* dependence: $X_j$ is important for predicting $Y$, given $X_{-j}$, i.e. $\mathbb{E}(Y \,|\, X) \neq \mathbb{E}(Y \,|\, X_{-j})$.



# Testing conditional independence

# Conditional independence testing is hard

We are interested in testing the null hypothesis $X \perp\!\!\!\perp Y \mid Z$ based on data.

- *Size*: Maximum probability of rejecting the null. We want this to be small.
- *Power*: Probability of rejecting when $X \not\perp\!\!\!\perp Y \mid Z$. We want this to be large.

# Conditional independence testing is hard

We are interested in testing the null hypothesis $X \perp\!\!\!\perp Y \mid Z$ based on data.

- *Size*: Maximum probability of rejecting the null. We want this to be small.
- *Power*: Probability of rejecting when $X \not\perp\!\!\!\perp Y \mid Z$. We want this to be large.

**Theorem**

*Suppose $Z$ is a continuous random variable. For any test,*

$$\textit{size} \geq \textit{power at any alternative.}$$

# Conditional independence testing is hard

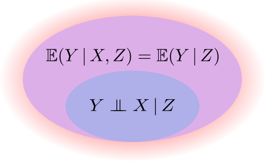
We are interested in testing the null hypothesis $X \perp\!\!\!\perp Y \mid Z$ based on data.

- *Size*: Maximum probability of rejecting the null. We want this to be small.
- *Power*: Probability of rejecting when $X \not\perp\!\!\!\perp Y \mid Z$. We want this to be large.

**Theorem**

*Suppose $Z$ is a continuous random variable. For any test,*

$$\textit{size} \geq \textit{power at any alternative.}$$



# Conditional independence testing is hard

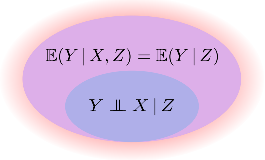
We are interested in testing the null hypothesis $X \perp\!\!\!\perp Y \mid Z$ based on data.

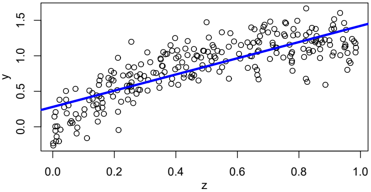
- *Size*: Maximum probability of rejecting the null. We want this to be small.
- *Power*: Probability of rejecting when $X \not\perp\!\!\!\perp Y \mid Z$. We want this to be large.

**Theorem**

*Suppose $Z$ is a continuous random variable. For any test,*

$$\textit{size} \geq \textit{power at any alternative.}$$




One way of restricting the null is via models...

# Conditional independence testing is hard

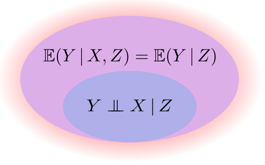
We are interested in testing the null hypothesis $X \perp\!\!\!\perp Y \mid Z$ based on data.

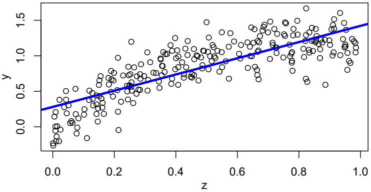
- *Size*: Maximum probability of rejecting the null. We want this to be small.
- *Power*: Probability of rejecting when $X \not\perp\!\!\!\perp Y \mid Z$. We want this to be large.

**Theorem**

*Suppose $Z$ is a continuous random variable. For any test,*

$$\textit{size} \geq \textit{power at any alternative.}$$




One way of restricting the null is via models...


...or we can attempt to leverage the predictive power of flexible regression methods.

# Generalised Covariance Measures

# Generalised Covariance Measure

We wish to test conditional independence with data $(X_i, Y_i, Z_i)_{i=1}^n$.

1. Using your favourite machine learning methods
   - Regress $(X_i)_{i=1}^n$ onto $(Z_i)_{i=1}^n$ to give fitted regression function $\hat{m}_{X|Z}$.
   - Regress $(Y_i)_{i=1}^n$ onto $(Z_i)_{i=1}^n$ to give fitted regression function $\hat{m}_{Y|Z}$.
2. Form the product of the residuals

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

3. Form a normalised empirical covariance of the $(L_i)_{i=1}^n$:

$$\mathrm{GCM} := \sqrt{n}\frac{\frac{1}{n}\sum_{i=1}^n L_i}{\{\frac{1}{n}\sum_{i=1}^n (L_i - \bar{L})^2\}^{1/2}}.$$

4. Compare GCM to quantiles of a standard normal to obtain $p$-value.

Key point: Validity of the test (i.e. control of type I error) relies primarily on the machine learning methods used being able to predict well.

# Generalised Covariance Measure

We wish to test conditional independence with data $(X_i, Y_i, Z_i)_{i=1}^n$.

1. Using your favourite machine learning methods
   - Regress $(X_i)_{i=1}^n$ onto $(Z_i)_{i=1}^n$ to give fitted regression function $\hat{m}_{X|Z}$.
   - Regress $(Y_i)_{i=1}^n$ onto $(Z_i)_{i=1}^n$ to give fitted regression function $\hat{m}_{Y|Z}$.
2. Form the product of the residuals

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

3. Form a normalised empirical covariance of the $(L_i)_{i=1}^n$:

$$\mathrm{GCM} := \sqrt{n}\frac{\frac{1}{n}\sum_{i=1}^n L_i}{\{\frac{1}{n}\sum_{i=1}^n (L_i - \bar{L})^2\}^{1/2}}.$$

4. Compare GCM to quantiles of a standard normal to obtain $p$-value.

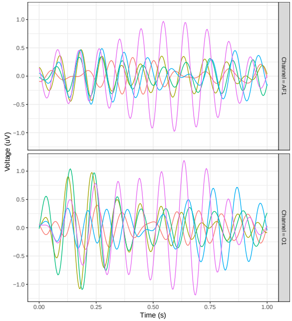
Key point: Validity of the test (i.e. control of type I error) relies primarily on the machine learning methods used being able to predict well.

- Related PCM test for conditional mean independence can be more powerful.
- Version of the GCM available for multivariate $X, Y$.

# Functional data

There are many settings where it is natural to consider the data observed as sampled values of functions.
E.g. Height or weight curves over time, yearly curves of rainfall or temperature, EEG data

Figure from Mayo Clinic

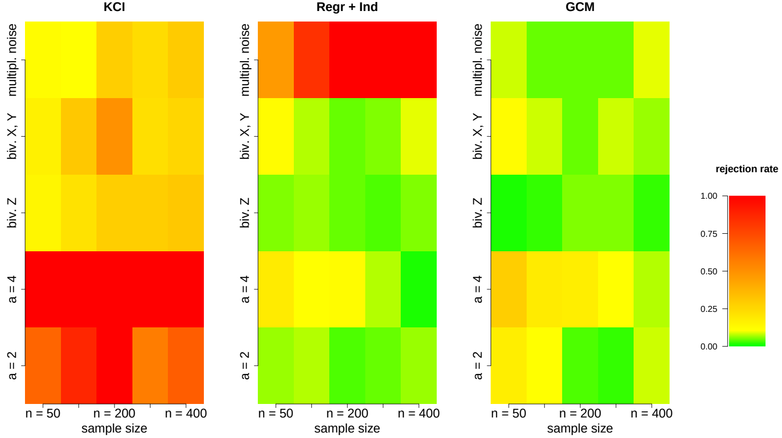



Zhang et al., 1995 and Ingber, 1997 study EEG data generated subjects with a genetic predisposition to alcoholism and a control group after being presented with a stimulus.

GHCM test available for testing conditional independence with functional data.

# Numerical experiments

# Comparison of GCM with other tests in null settings

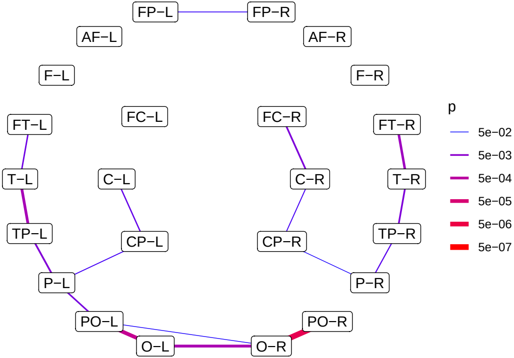



Comparison with KCI: (Zhang et al., 2011),

Reg + Ind: (e.g. Ramsey, 2014)

# EEG data from alcoholism study

- EEG data on 77 subjects.
- Form conditional independence graph by testing conditional independencies using the GHCM.
- Functional regressions performed using `FDboost`.



# Summary

- Conditional independence testing is important but hard.
- Some domain knowledge is required in order to select an appropriate test for a particular setting.
- The GCM / GHCM/ PCM frameworks give conditional (mean) independence tests whose validity depends primarily on predictive properties of machine learning methods.
- Papers:


Shah, R. D. and Peters, J. (2020) The hardness of conditional independence and the generalised covariance measure. *Annals of Statistics*.

Lundborg, A, Shah, R. D. and Peters, J. (2022) Conditional Independence Testing in Hilbert Spaces with Applications to Functional Data Analysis. *JRSSB*.

Lundborg, A. R., Kim, I., Shah, R. D. and Samworth, R. J. (2022) The Projected Covariance Measure for assumption-lean variable significance testing. *arXiv Preprint*.


*Thank you for listening.*