Controlling the false positive rate of Granger causality tests in fMRI data

Oliver (and Leo, Ben, Mac, Joe)

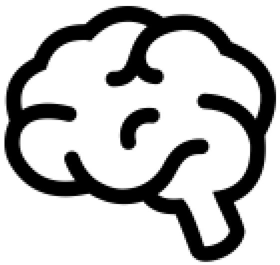
THE UNIVERSITY OF SYDNEY

Increased false positive rates

Ridiculous question: Does one patient’s fMRI data influence another’s?

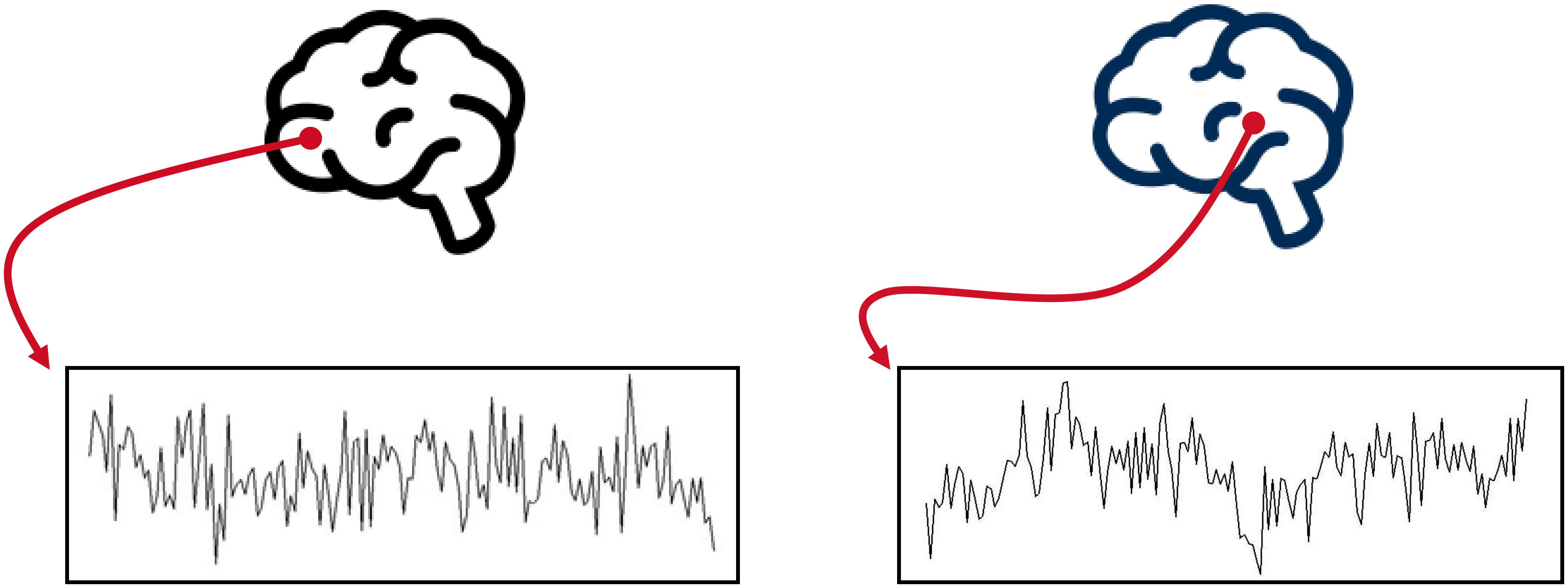
Increased false positive rates

Ridiculous question: Does one patient’s fMRI data influence another’s?



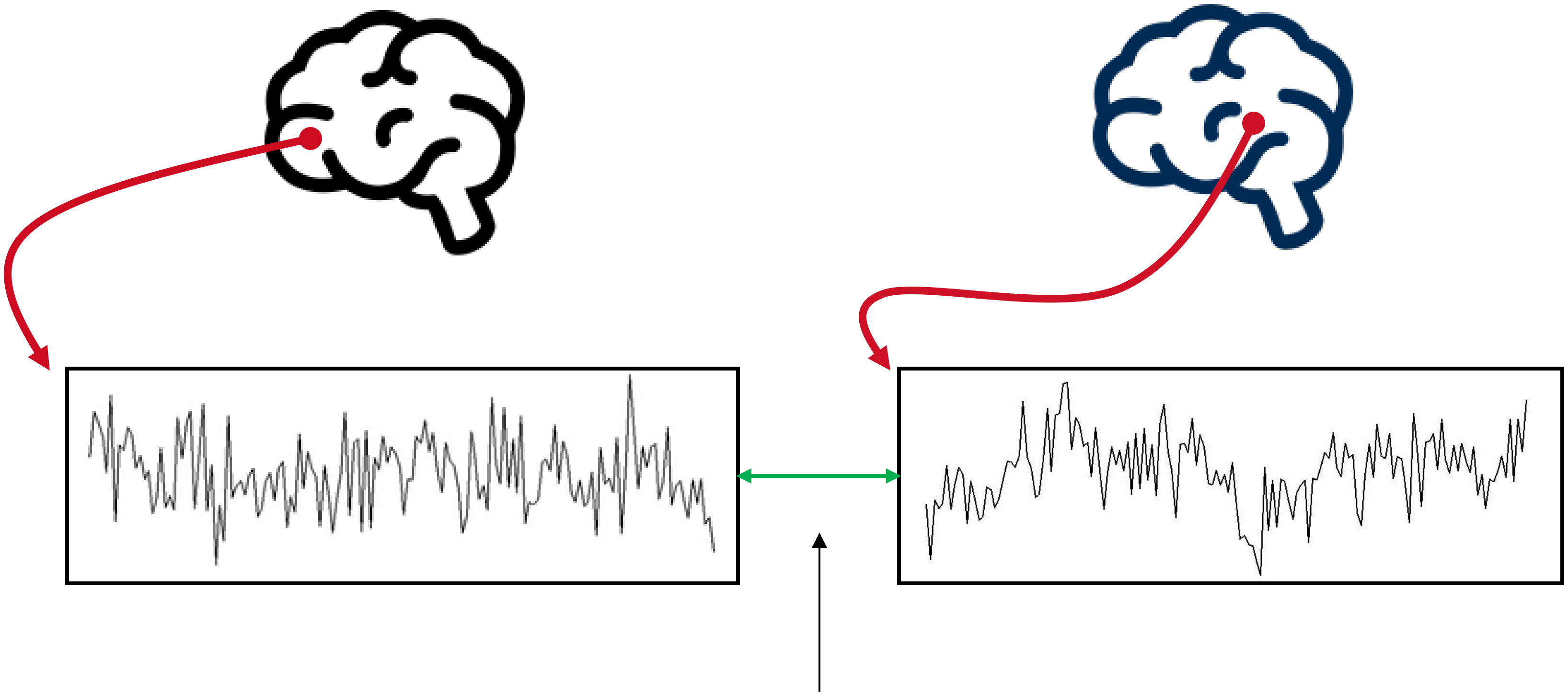
Increased false positive rates

Ridiculous question: Does one patient's fMRI data influence another's?



Increased false positive rates

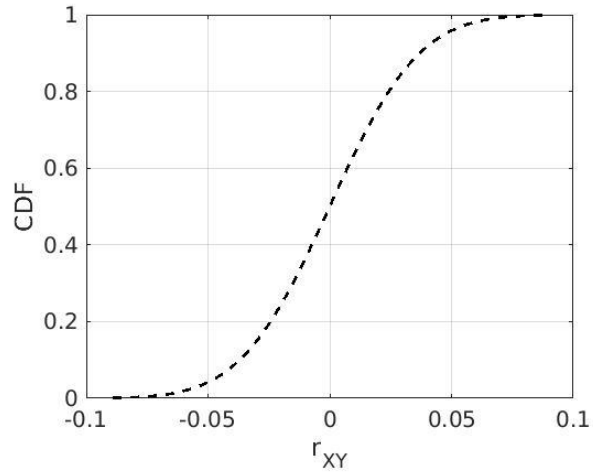
Ridiculous question: Does one patient's fMRI data influence another's?



correlation, mutual information, Granger causality, etc.

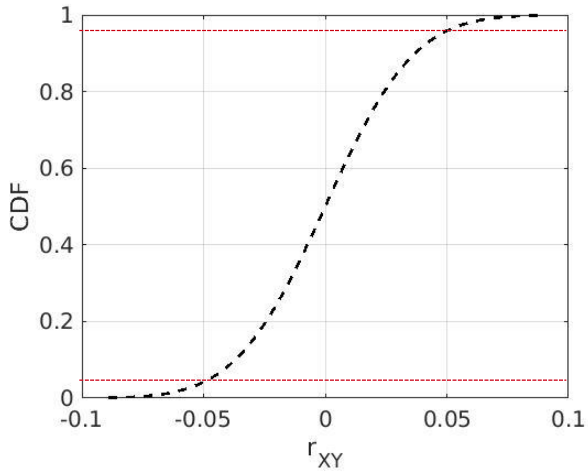
Increased false positive rates

Test correlation against a t distribution…



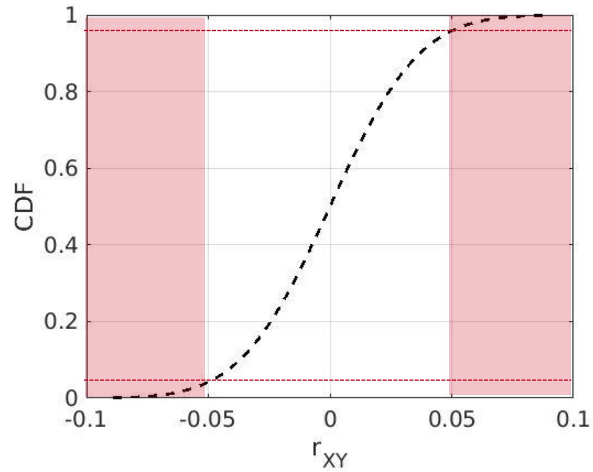
Increased false positive rates

…at a nominal p-value of 5%



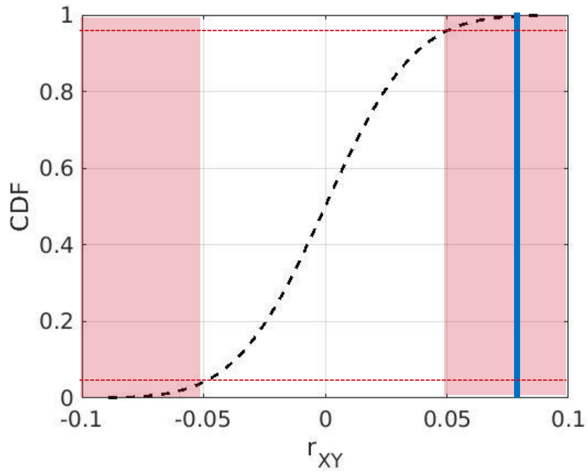
Increased false positive rates

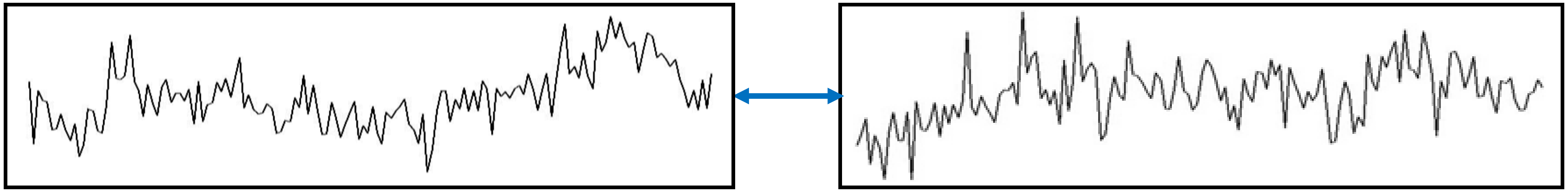
…at a nominal p-value of 5%



Increased false positive rates

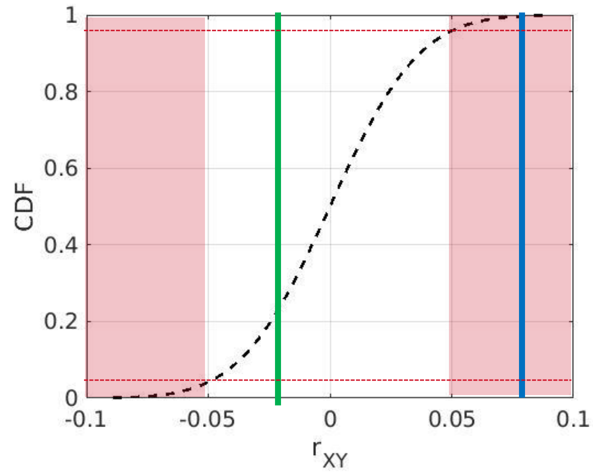
Answer: 1/1 tests indicate true (100%)…

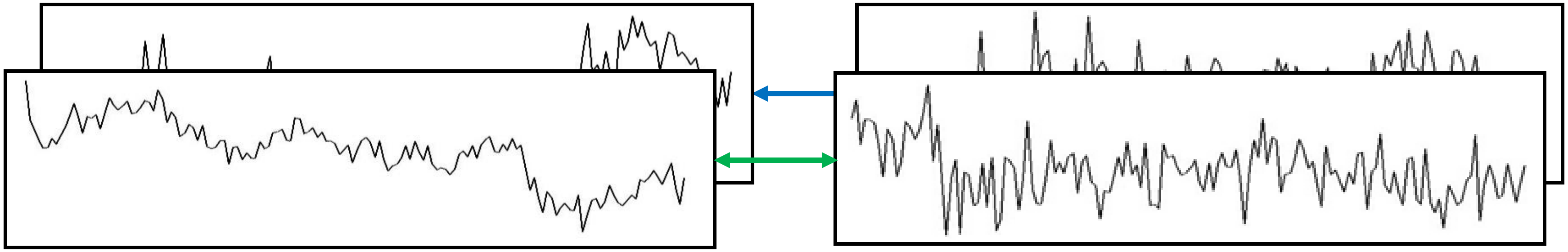




Increased false positive rates

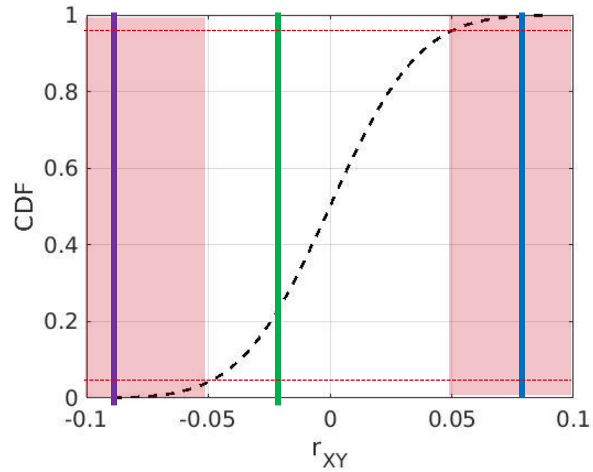
Answer: 1/2 tests indicate true (50%)…

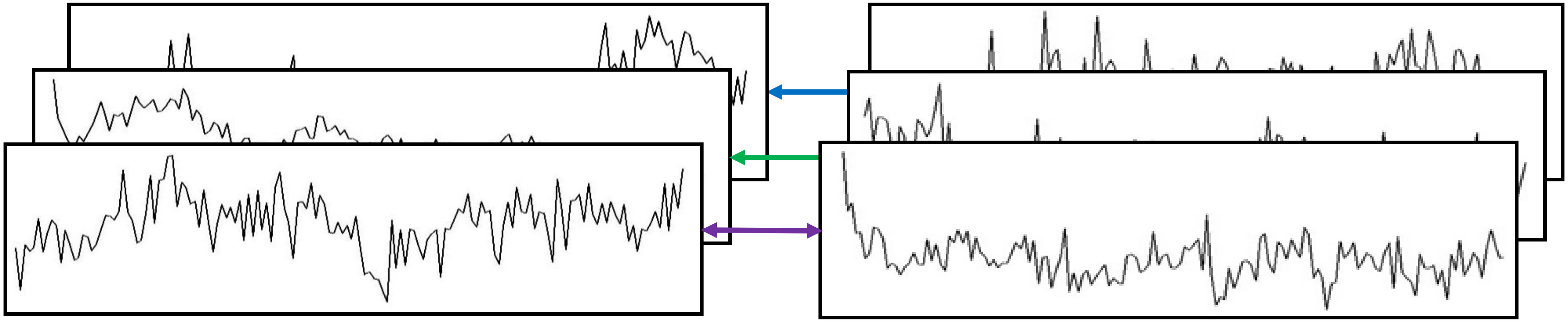




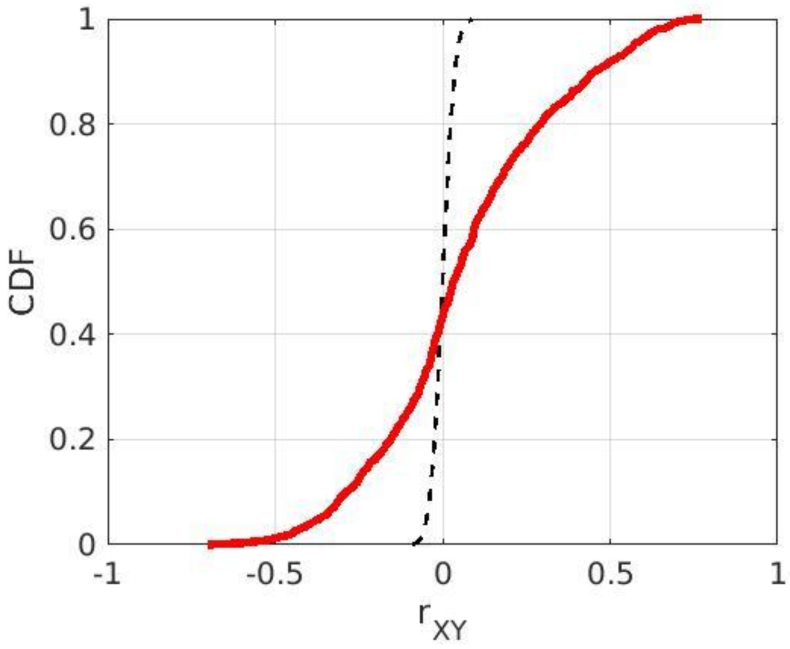
Increased false positive rates

Answer: 2/3 tests indicate true (67%)…



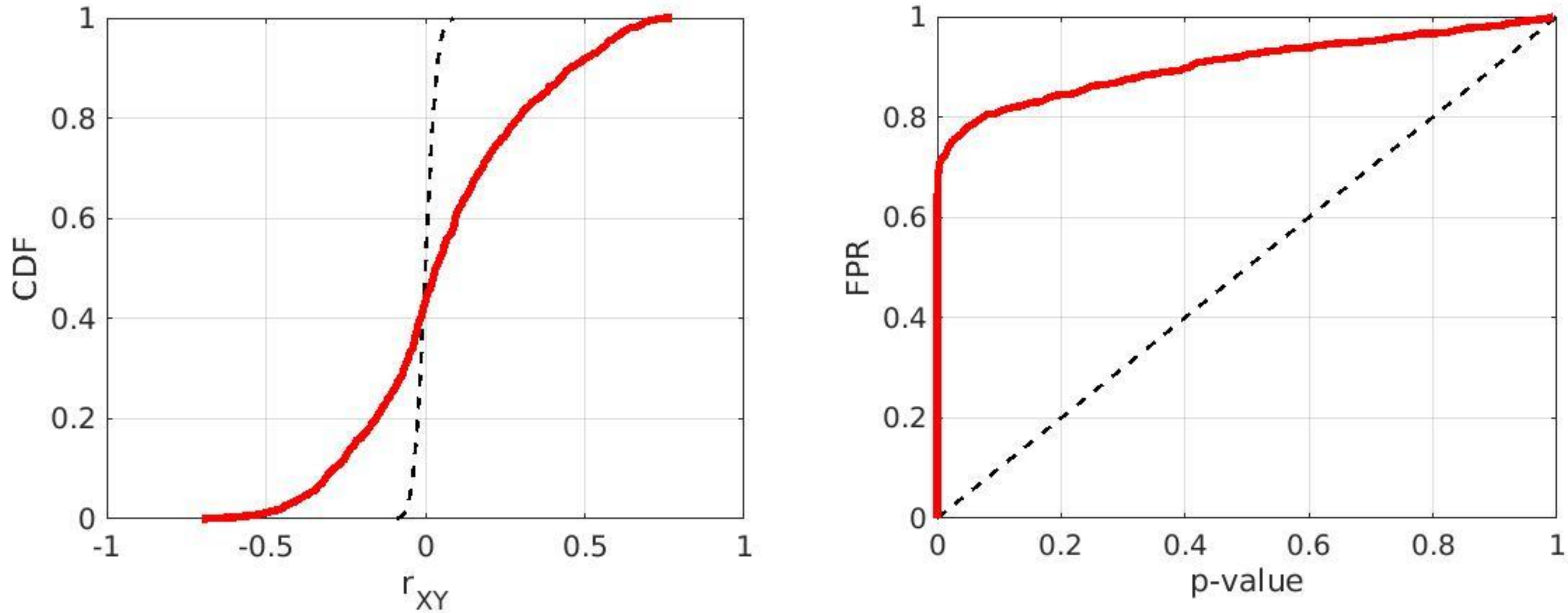


Answer: 781/1000 tests indicate true (78%)



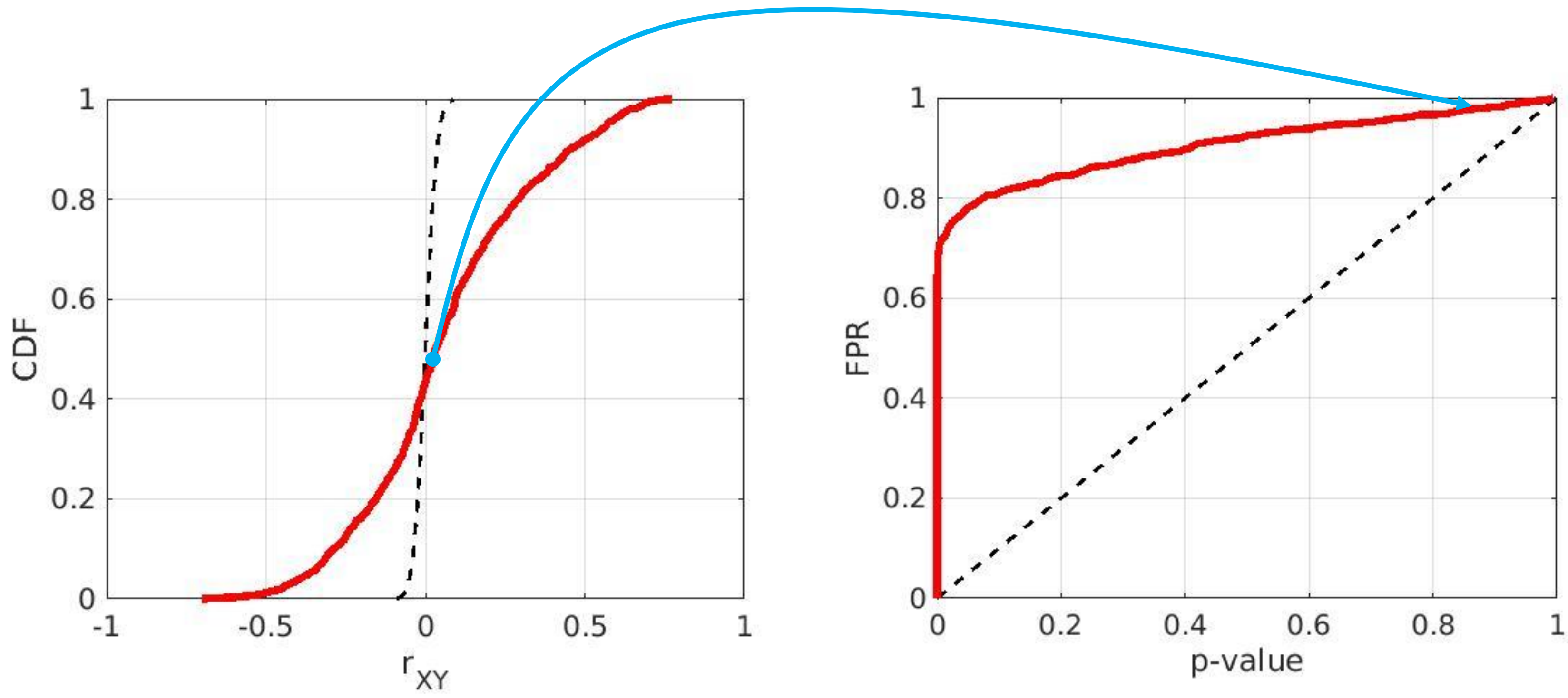
Increased false positive rates

Answer: 781/1000 tests indicate true (78%)



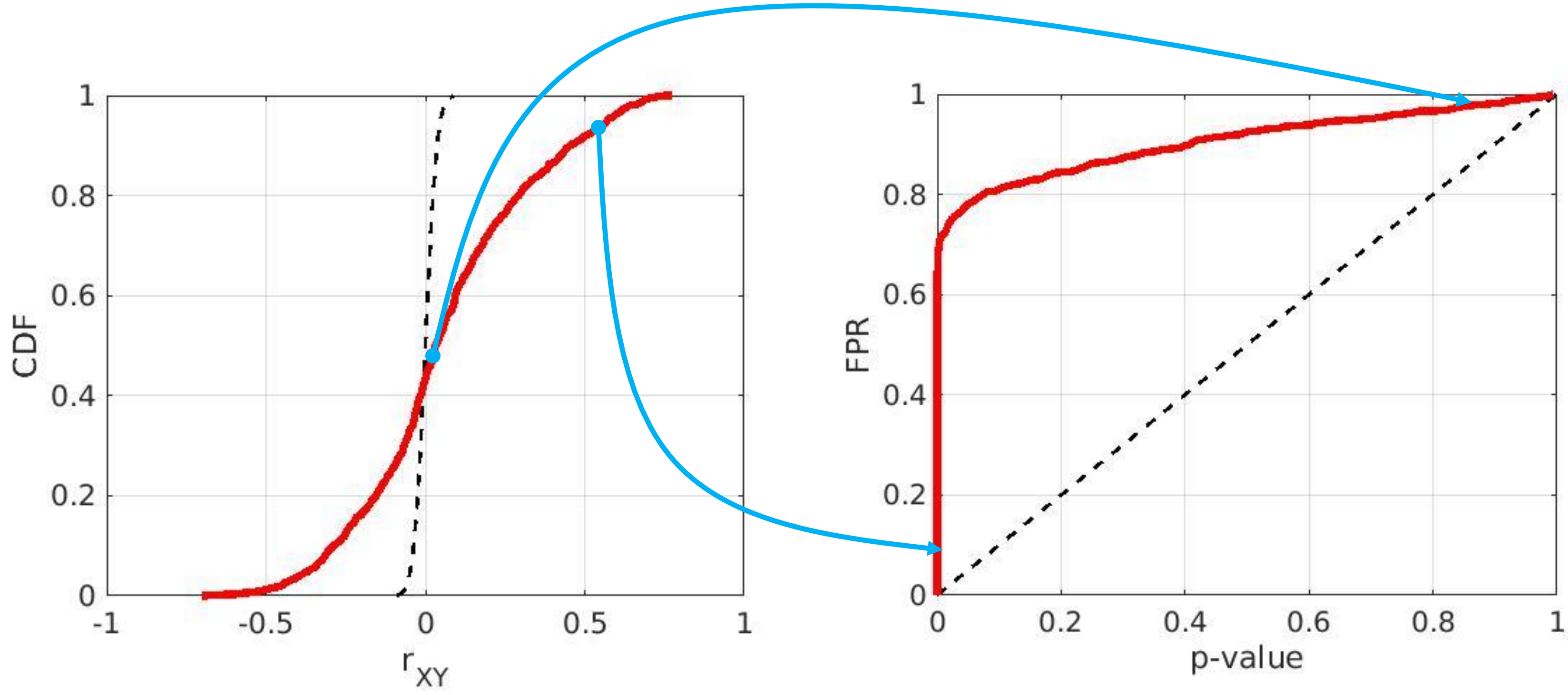
Increased false positive rates

Answer: 781/1000 tests indicate true (78%)



Increased false positive rates

Answer: 781/1000 tests indicate true (78%)

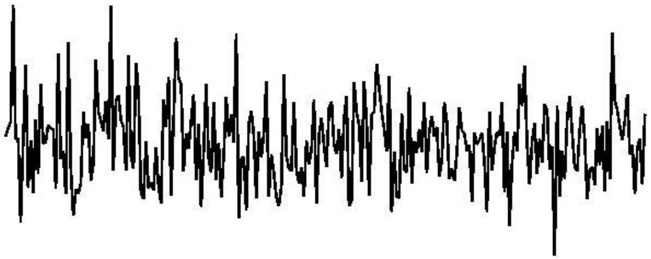


Increased false positive rates

- Pearson correlation assumptions:
  - **Gaussian** (innovations are drawn from bivariate normal dist.)
  - **IID** (no autocorrelation)
  - **Linear**

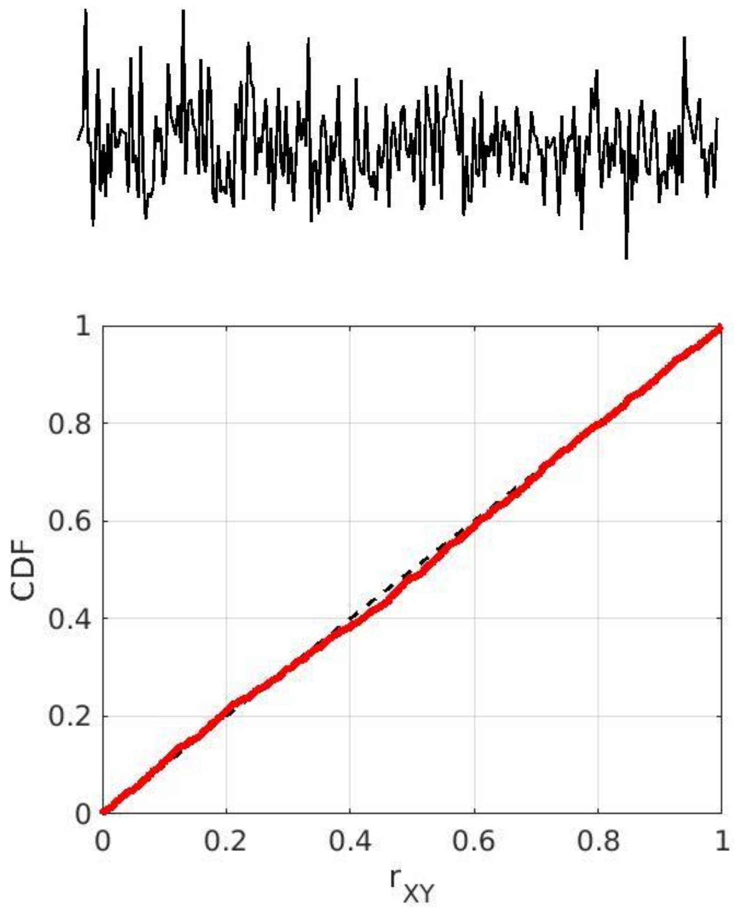
Increased false positive rates

Test: Gaussian variables using synthetic data (N=256)



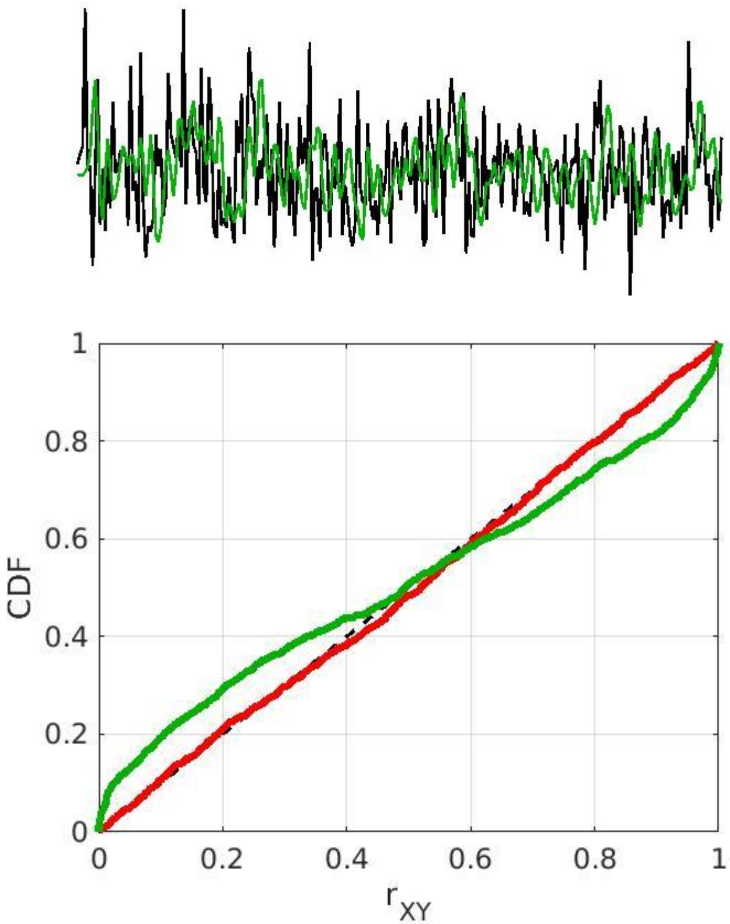
Increased false positive rates

Test: IID Gaussian variables using synthetic data (N=256)



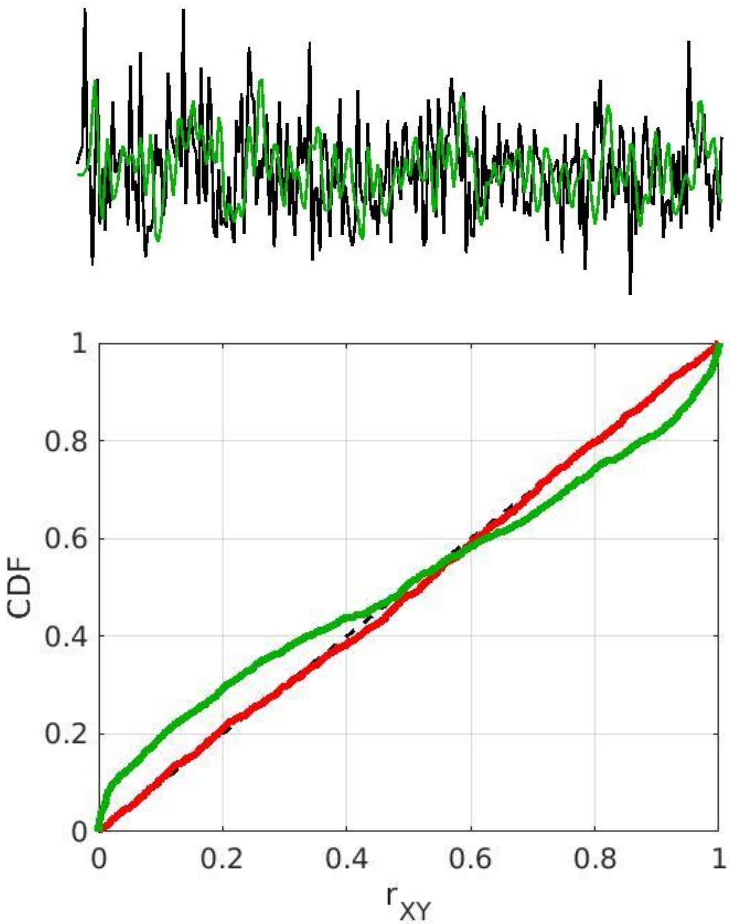
Increased false positive rates

Test: correlated Gaussian variables using synthetic data (N=256)



Increased false positive rates

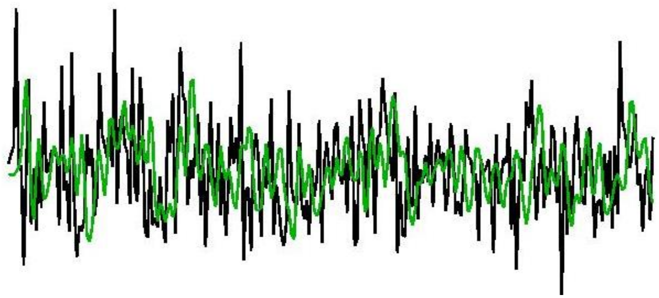
Test: correlated Gaussian variables using synthetic data (N=256)



- Weakly stationary processes are fully defined by a constant:
  - Mean
  - Autocorrelation

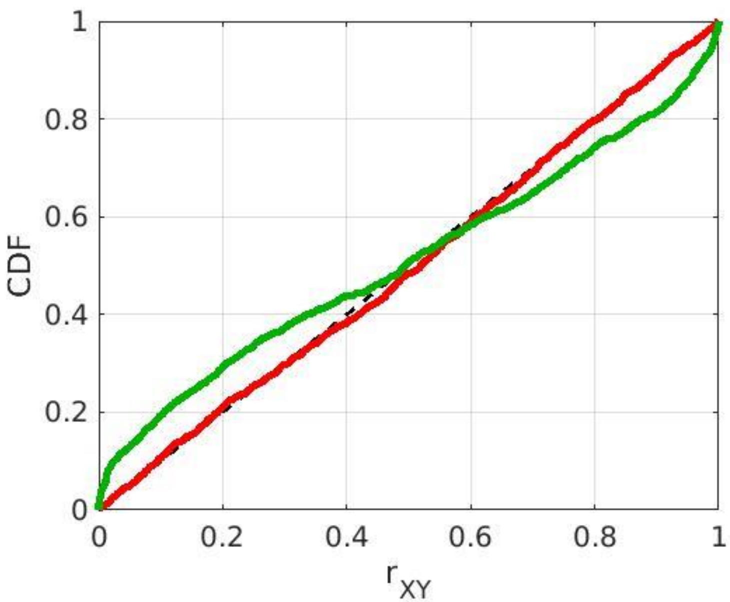
Increased false positive rates

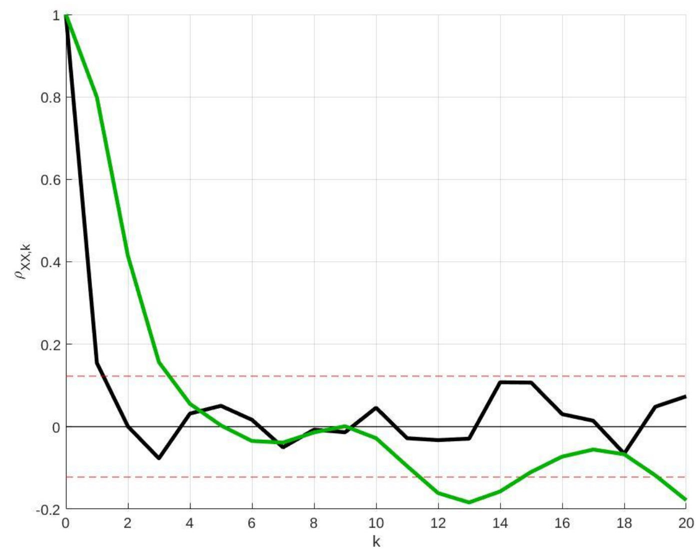
Test: correlated Gaussian variables using synthetic data (N=256)



› Weakly stationary processes are fully defined by a constant:

- Mean
- Autocorrelation





Increased false positive rates

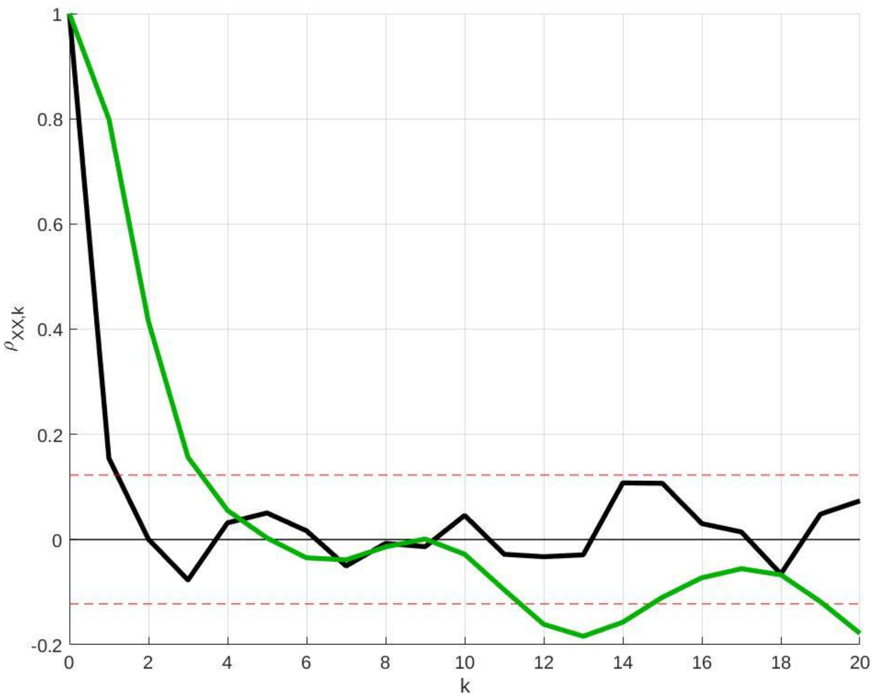
- Higher autocorrelation increases the FPR of sample correlation
  - This has been known since Yule/Pearson's work in the early 20th century

- How do we fix it?
  - **Bartlett's formula**
    - Correct the sample size based on autocorrelation of univariate signals
    - Does the process pass the t-test with effective sample size?
  - **Granger causality**
    - Define process through autocovariance
    - Does adding in another process improve predictability?

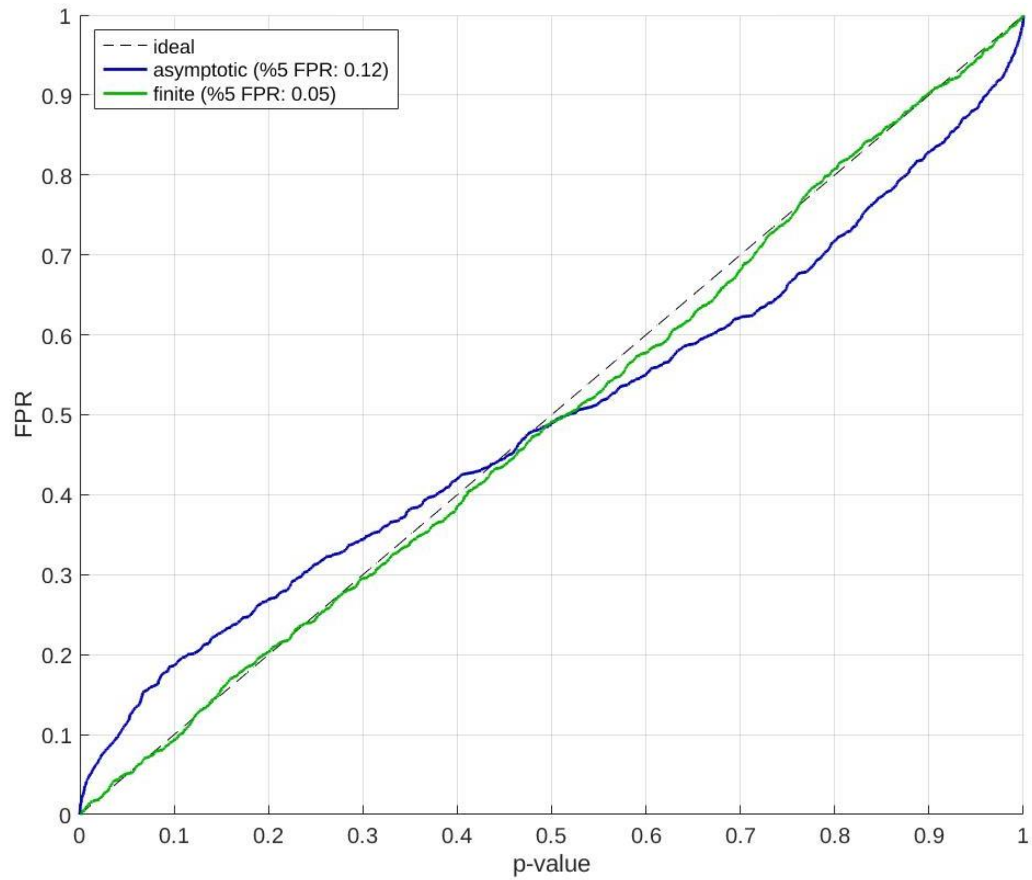
Bartlett's formula

- Testing Pearson correlation:
  - On WB

- Testing Pearson correlation:
  - On WB
- Under autocorrelation:
  - On WB



Bartlett’s formula



- Asymptotic distribution is **inefficient**
- Bartlett corrected values correct to nominal FPR

Granger causality

- Build a target AR process
  - [On WB]

- (Steal joe's figure for TE)

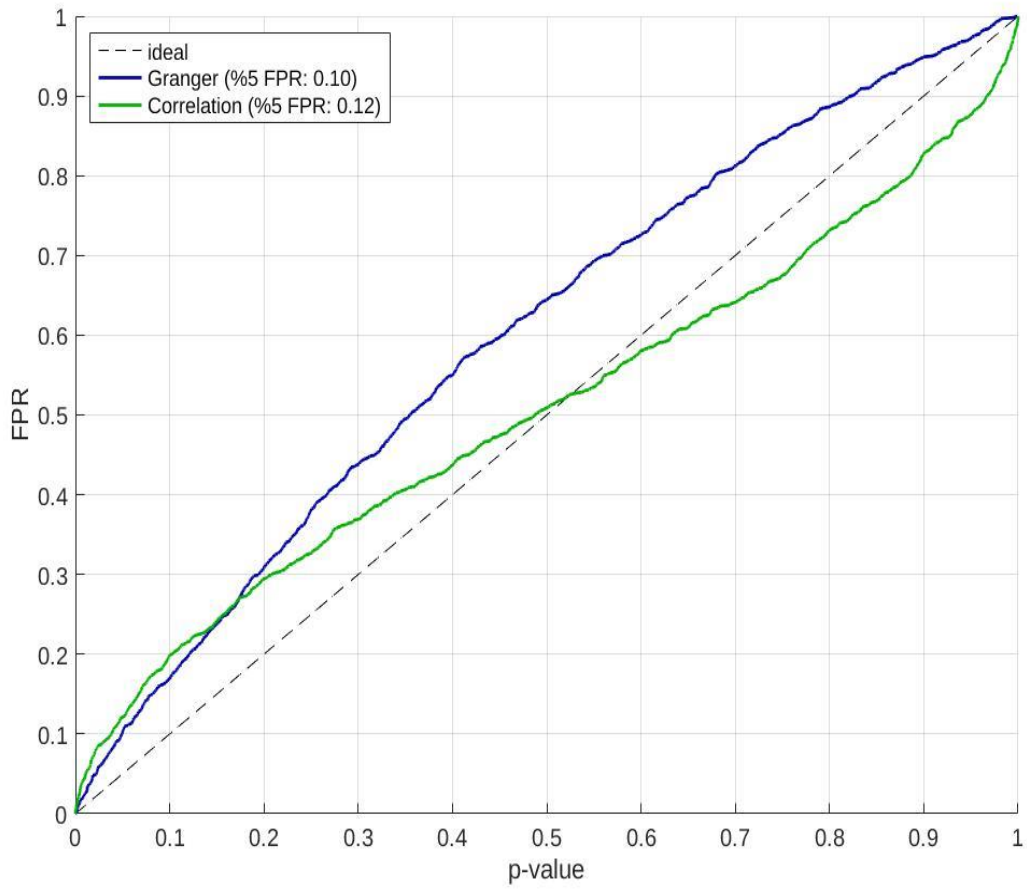
Granger causality

- Build a target AR process
  - [On WB]
- Include another (source) process
  - [On WB]
- (Steal joe's figure for TE)

Granger causality

- Build a target AR process
  - [On WB]
- Include another (source) process
  - [On WB]
- Test using Wilk’s theorem:
  - [On WB]
- (Steal joe’s figure for TE)

Granger causality



- Asymptotic distribution is **biased**
- Slightly lower FPR than correlation

THE UNIVERSITY OF SYDNEY

Bartlett's formula and Granger causality

- Bartlett's formula controls FPR for univariate processes
  - Multivariate generalization is known as canonical correlations
    - These are also inefficient under serial correlation

- Granger causality reduces but does not control FPR
  - Naturally extends to multiple time series
    - FPR increases for:
      - more dimensions, or
      - higher order AR/filtering

Sampling distribution of Granger causality

- How do we control for the FPR of Granger?

- Key insight: GC can be expressed in terms of correlation
  - Granger causality is equivalent to **conditional mutual information** (Barnett et al., 2009)
  - We show CMI can be expressed as a squared **partial correlation**
  - Through chain rule, multivariate measures are sums of squared PC
  - Bartlett’s formula can be used for the sampling distributions

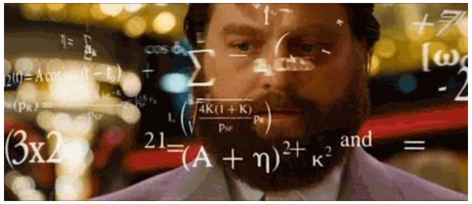
Sampling distribution of log-likelihoods

- Many dependence measures/tests have their sampling distributions derived from Wilk's theorem
  - Theorem assumes IID variables (invalid for stochastic processes)

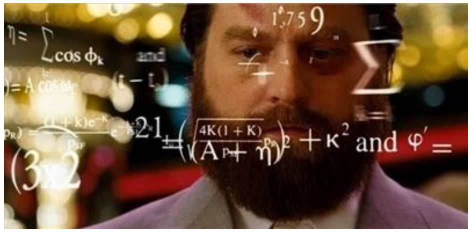
- No known finite-sample distribution for stochastic processes for:
  - Mutual information
  - Conditional mutual information
  - Transfer entropy/Granger causality

- We provide

Mutual information as squared correlation



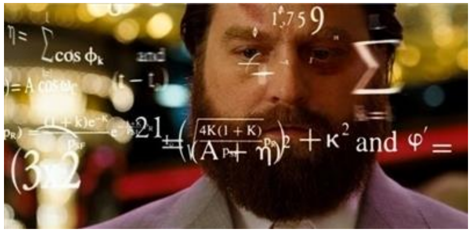
Mutual information as squared correlation



- MI for univariate Gaussian processes:

$$I_{X;Y} = -\frac{1}{2}\log\left(\frac{|\boldsymbol{K}_{XY}|}{|\boldsymbol{K}_X||\boldsymbol{K}_Y|}\right)$$

Mutual information as squared correlation



- MI for univariate Gaussian processes:

$$I_{X;Y} = -\frac{1}{2}\log\left(\frac{|\boldsymbol{K}_{XY}|}{|\boldsymbol{K}_X||\boldsymbol{K}_Y|}\right)$$

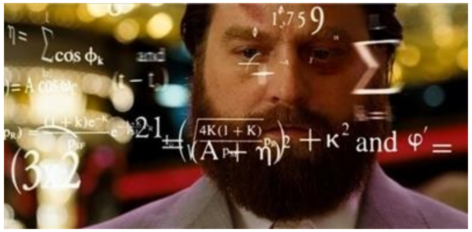
- Covariance matrices

$$K_X = \mathrm{Var}(X)$$

$$K_Y = \mathrm{Var}(Y)$$

$$\boldsymbol{K}_{XY} = \begin{bmatrix} \mathrm{Var}(X) & \mathrm{Cov}(X,Y) \\ \mathrm{Cov}(Y,X) & \mathrm{Var}(Y) \end{bmatrix}$$

Mutual information as squared correlation



- MI for univariate Gaussian processes:

$$I_{X;Y} = -\frac{1}{2}\log\left(\frac{|\boldsymbol{K}_{XY}|}{|\boldsymbol{K}_X||\boldsymbol{K}_Y|}\right)$$

- Covariance matrices

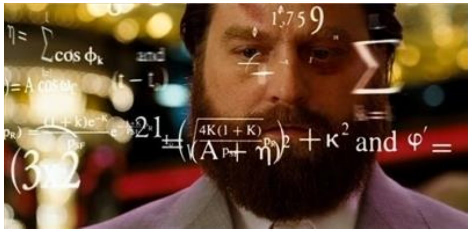
$$K_X = \mathrm{Var}(X) \qquad \boldsymbol{K}_{XY} = \begin{bmatrix} \mathrm{Var}(X) & \mathrm{Cov}(X,Y) \\ \mathrm{Cov}(Y,X) & \mathrm{Var}(Y) \end{bmatrix}$$

$$K_Y = \mathrm{Var}(Y)$$

- Taking the determinants gives us:

$$I_{X;Y} = -1/2\,\log(1-\rho_{XY}^2)$$

Mutual information as squared correlation



- MI for univariate Gaussian processes:

$$I_{X;Y} = -\frac{1}{2}\log\left(\frac{|\boldsymbol{K}_{XY}|}{|\boldsymbol{K}_X||\boldsymbol{K}_Y|}\right)$$

- Covariance matrices

$$K_X = \mathrm{Var}(X) \qquad \boldsymbol{K}_{XY} = \begin{bmatrix} \mathrm{Var}(X) & \mathrm{Cov}(X,Y) \\ \mathrm{Cov}(Y,X) & \mathrm{Var}(Y) \end{bmatrix}$$

$$K_Y = \mathrm{Var}(Y)$$

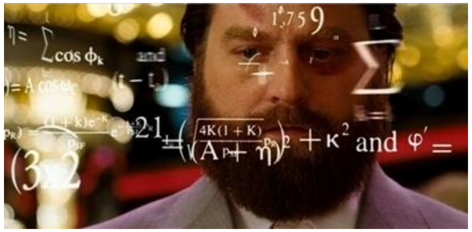
- Taking the determinants gives us:

$$I_{X;Y} = -1/2\,\log(1-\rho_{XY}^2)$$

- First-order Taylor expansion:

$$I_{X;Y} \approx 1/2\,\rho_{XY}^2$$

Mutual information as squared correlation



› MI for univariate Gaussian processes:

$$I_{X;Y} = -\frac{1}{2}\log\left(\frac{|\boldsymbol{K}_{XY}|}{|\boldsymbol{K}_X||\boldsymbol{K}_Y|}\right)$$

› Covariance matrices

$$K_X = \mathrm{Var}(X)$$

$$K_Y = \mathrm{Var}(Y)$$

$$\boldsymbol{K}_{XY} = \begin{bmatrix} \mathrm{Var}(X) & \mathrm{Cov}(X,Y) \\ \mathrm{Cov}(Y,X) & \mathrm{Var}(Y) \end{bmatrix}$$

› Taking the determinants gives us:

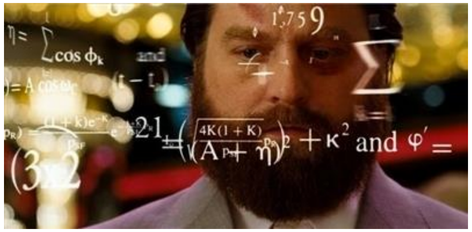
$$I_{X;Y} = -1/2\,\log(1-\rho^2_{XY})$$

› First-order Taylor expansion:

$$\boxed{I_{X;Y} \approx 1/2\,\rho^2_{XY}}$$

Approximate sampling distribution

Mutual information as squared correlation



- MI for univariate Gaussian processes:

$$I_{X;Y} = -\frac{1}{2}\log\left(\frac{|\boldsymbol{K}_{XY}|}{|\boldsymbol{K}_X||\boldsymbol{K}_Y|}\right)$$

- Covariance matrices

$$K_X = \mathrm{Var}(X) \qquad \boldsymbol{K}_{XY} = \begin{bmatrix} \mathrm{Var}(X) & \mathrm{Cov}(X,Y) \\ \mathrm{Cov}(Y,X) & \mathrm{Var}(Y) \end{bmatrix}$$

$$K_Y = \mathrm{Var}(Y)$$

- Taking the determinants gives us:

$$\boxed{I_{X;Y} = -1/2\,\log(1-\rho^2_{XY})}$$

Exact sampling distribution

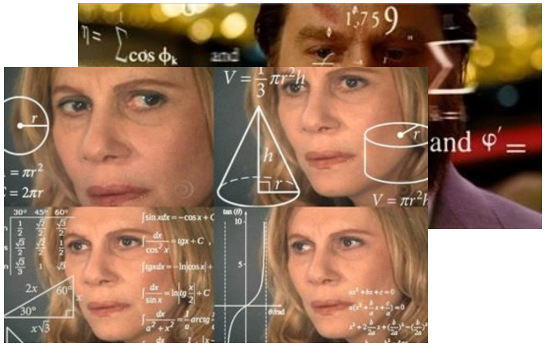
- First-order Taylor expansion:

$$\boxed{I_{X;Y} \approx 1/2\,\rho^2_{XY}}$$

Approximate sampling distribution

THE UNIVERSITY OF SYDNEY

Conditional mutual information as squared partial correlation



- CMI for univariate Gaussian processes:

$$I_{X;Y|\boldsymbol{W}} = -\frac{1}{2}\log\left(\frac{|\boldsymbol{K}_{XY|\boldsymbol{W}}|}{|\boldsymbol{K}_{X|\boldsymbol{W}}||\boldsymbol{K}_{Y|\boldsymbol{W}}|}\right)$$

- Covariance matrices

$$K_{X|\boldsymbol{W}} = \mathrm{Var}(X|\boldsymbol{W}) \qquad \boldsymbol{K}_{XY|\boldsymbol{W}} = \begin{bmatrix} \mathrm{Var}(X|\boldsymbol{W}) & \mathrm{Cov}(X,Y|\boldsymbol{W}) \\ \mathrm{Cov}(Y,X|\boldsymbol{W}) & \mathrm{Var}(Y|\boldsymbol{W}) \end{bmatrix}$$

$$K_{Y|\boldsymbol{W}} = \mathrm{Var}(Y|\boldsymbol{W})$$

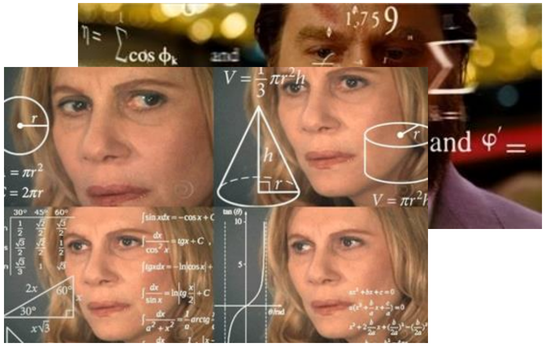
- Taking the determinants gives us:

$$I_{X;Y|\boldsymbol{W}} = -1/2\,\log(1-\rho^2_{XY\cdot\boldsymbol{W}})$$

- First-order Taylor expansion:

$$I_{X;Y|\boldsymbol{W}} \approx 1/2\,\rho^2_{XY\cdot\boldsymbol{W}}$$

Conditional mutual information as squared partial correlation



- CMI for univariate Gaussian processes:

$$I_{X;Y|\boldsymbol{W}} = -\frac{1}{2}\log\left(\frac{|\boldsymbol{K}_{XY|\boldsymbol{W}}|}{|\boldsymbol{K}_{X|\boldsymbol{W}}||\boldsymbol{K}_{Y|\boldsymbol{W}}|}\right)$$

- Covariance matrices

$$K_{X|\boldsymbol{W}} = \mathrm{Var}(X|\boldsymbol{W}) \qquad \boldsymbol{K}_{XY|\boldsymbol{W}} = \begin{bmatrix} \mathrm{Var}(X|\boldsymbol{W}) & \mathrm{Cov}(X,Y|\boldsymbol{W}) \\ \mathrm{Cov}(Y,X|\boldsymbol{W}) & \mathrm{Var}(Y|\boldsymbol{W}) \end{bmatrix}$$

$$K_{Y|\boldsymbol{W}} = \mathrm{Var}(Y|\boldsymbol{W})$$

- Taking the determinants gives us:

$$\boxed{I_{X;Y|\boldsymbol{W}} = -1/2\,\log(1-\rho^2_{XY\cdot\boldsymbol{W}})}$$

Exact sampling distribution

- First-order Taylor expansion:

$$\boxed{I_{X;Y|\boldsymbol{W}} \approx 1/2\,\rho^2_{XY\cdot\boldsymbol{W}}}$$

Approximate sampling distribution

What does this mean for GC (and others)?

- Partial correlation is the correlation between two residuals:

$$\rho_{XY\cdot\boldsymbol{W}} = \rho_{e_X e_Y}$$

- Computed w.r.t the conditional:

$$e_X = X - \boldsymbol{W}\beta^*_X$$
$$e_Y = Y - \boldsymbol{W}\beta^*_Y$$

- Approximate sampling distribution:

$$I_{X;Y|\boldsymbol{W}} \approx 1/2\,\rho^2_{XY\cdot\boldsymbol{W}}$$

- Exact sampling distribution:

Change these to this

$$I_{X;Y|\boldsymbol{W}} = -1/2\,\log(1-\rho^2_{XY\cdot\boldsymbol{W}})$$

Sampling distribution of partial correlation and CMI

- Typically, sampling distribution of MI/CMI/Granger assumed chi-squared distribution
  - From Wilk's theorem.

- Partial correlation follows same distribution as correlation with less *effective samples* based on the number of conditionals
  - And CMI is just partial correlation squared

- Squaring t-distribution gives an F-distribution
  - Now we know the exact distribution of univariate CMI
  - What about multivariate measures?

Decomposing the multivariate measures

› Use the chain rule to decompose MV MI

$$
\begin{aligned}
I_{\boldsymbol{X};\boldsymbol{Y}} &= \sum_{i,j} I_{X_i;Y_j|\{\boldsymbol{X}_{1:i-1},\boldsymbol{Y}_{1:j-1}\}} \\
&= -1/2 \sum_{i,j} \log\left(1-\rho^2_{X_iY_j\cdot\{\boldsymbol{X}_{1:i-1},\boldsymbol{Y}_{1:j-1}\}}\right) \\
&\approx 1/2 \sum_{i,j} \rho^2_{X_iY_j\cdot\{\boldsymbol{X}_{1:i-1},\boldsymbol{Y}_{1:j-1}\}},
\end{aligned}
$$

Decomposing the multivariate measures

› Use the chain rule to decompose MV MI

$$
\begin{aligned}
I_{\boldsymbol{X};\boldsymbol{Y}} &= \sum_{i,j} I_{X_i;Y_j|\{\boldsymbol{X}_{1:i-1},\boldsymbol{Y}_{1:j-1}\}} \\
&= -1/2 \sum_{i,j} \log\left(1 - \rho^2_{X_i Y_j \cdot \{\boldsymbol{X}_{1:i-1},\boldsymbol{Y}_{1:j-1}\}}\right) \\
&\approx 1/2 \sum_{i,j} \boxed{\rho^2_{X_i Y_j \cdot \{\boldsymbol{X}_{1:i-1},\boldsymbol{Y}_{1:j-1}\}}}
\end{aligned}
$$

F distribution

Decomposing the multivariate measures

- …and MV CMI

$$
\begin{aligned}
I_{\boldsymbol{X};\boldsymbol{Y}|\boldsymbol{W}} &= \sum_{i,j} I_{X_i;Y_j|\{\boldsymbol{X}_{1:i-1},\boldsymbol{Y}_{1:j-1},\boldsymbol{W}\}} \\
&= -1/2 \sum_{i,j} \log\left(1-\rho^2_{X_iY_j\cdot\{\boldsymbol{X}_{1:i-1},\boldsymbol{Y}_{1:j-1},\boldsymbol{W}\}}\right) \\
&\approx 1/2 \sum_{i,j} \rho^2_{X_iY_j\cdot\{\boldsymbol{X}_{1:i-1},\boldsymbol{Y}_{1:j-1},\boldsymbol{W}\}}
\end{aligned}
$$

Decomposing the multivariate measures

- …and MV CMI

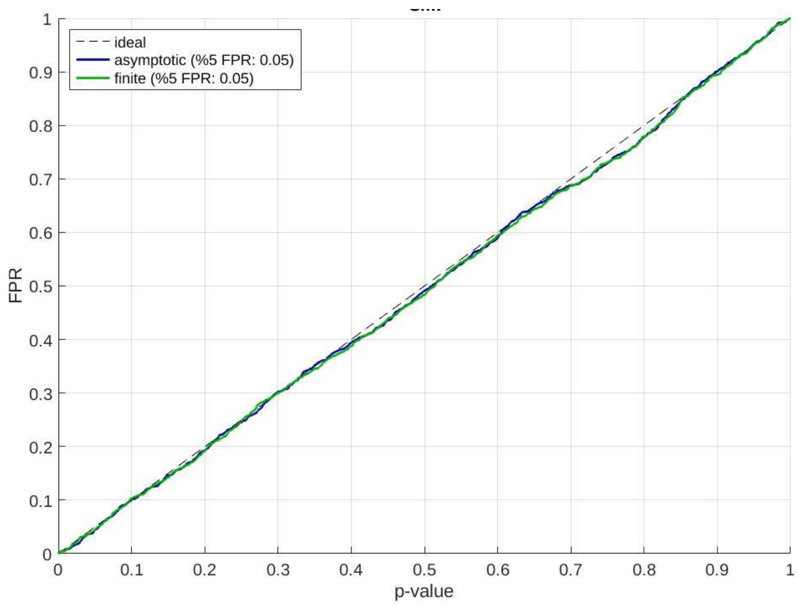
$$
\begin{aligned}
I_{\boldsymbol{X};\boldsymbol{Y}|\boldsymbol{W}} &= \sum_{i,j} I_{X_i;Y_j|\{\boldsymbol{X}_{1:i-1},\boldsymbol{Y}_{1:j-1},\boldsymbol{W}\}} \\
&= -1/2 \sum_{i,j} \log\left(1 - \rho^2_{X_i Y_j \cdot \{\boldsymbol{X}_{1:i-1},\boldsymbol{Y}_{1:j-1},\boldsymbol{W}\}}\right) \\
&\approx 1/2 \sum_{i,j} \boxed{\rho^2_{X_i Y_j \cdot \{\boldsymbol{X}_{1:i-1},\boldsymbol{Y}_{1:j-1},\boldsymbol{W}\}}}
\end{aligned}
$$

F distribution

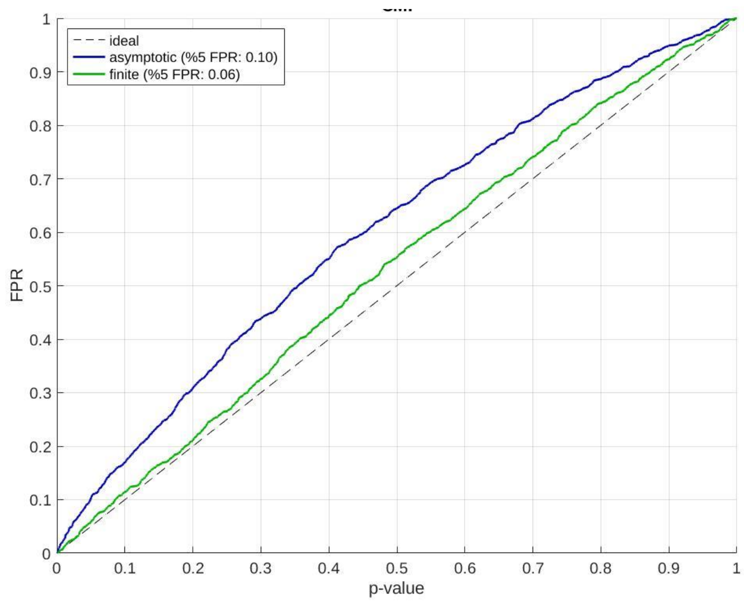
What does it all mean?

Controlling the FPR of GC

IID

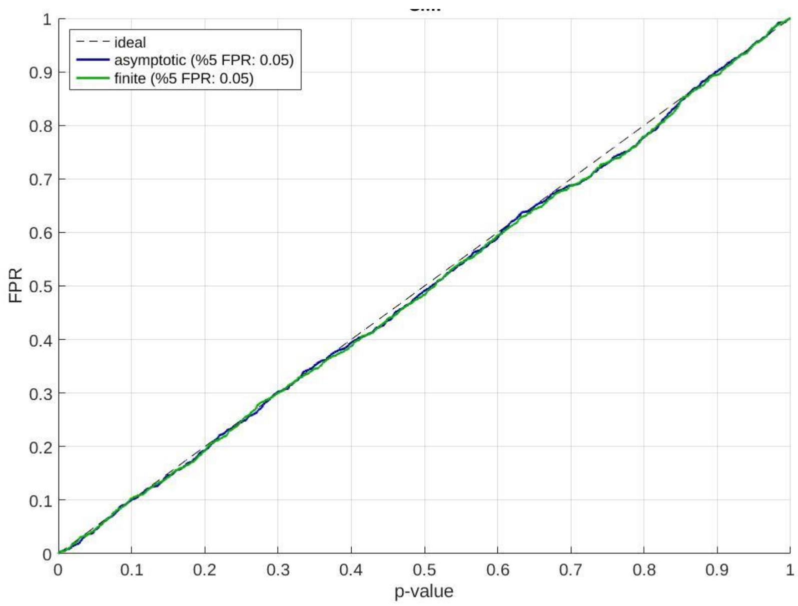


~8th Order AR
(FIR filter)

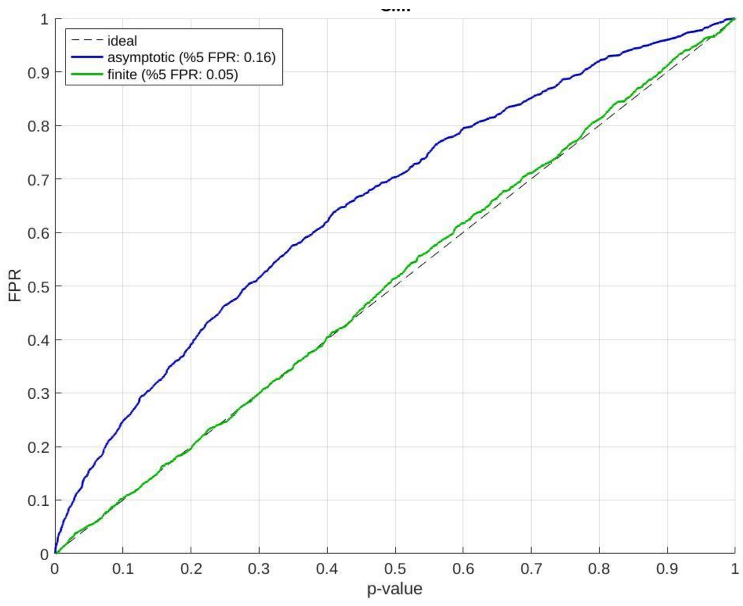


Controlling the FPR of GC

IID



~16th Order AR
(IIR filter)



Take home messages

- Important dependence measures exhibit bias for autocorrelated Gaussian processes
- These measures can be represented as sums of squared partial correlations
- This representation allows us to derive the sampling distribution
- Before our work, these distributions were only valid asymptotically

Take home messages

- Our empirical results confirm the asymptotic sample distribution yields higher FPs for:
  - Higher order AR/filters
  - Higher dimensional processes

- Our sampling distribution controls the FPR of (univariate and multivariate) measures of dependence:
  - Mutual information
  - Conditional mutual information
  - Granger causality

THE UNIVERSITY OF SYDNEY

Thanks!