10èmes Journées Francophones sur les Réseaux Bayésiens et les Modèles Graphiques Probabilistes,
JFRB-2021, 11-12.10.2021

Causal Identification with Additive Noise Models: Quantifying the Effect of Noise

Benjamin KAP, PhD Marharyta ALEKSANDROVA, Prof. Thomas ENGEL

SECAN-Lab

UNIVERSITÉ DU LUXEMBOURG

Outline

Introduction to Causal Discovery and state-of-the-art

RESIT

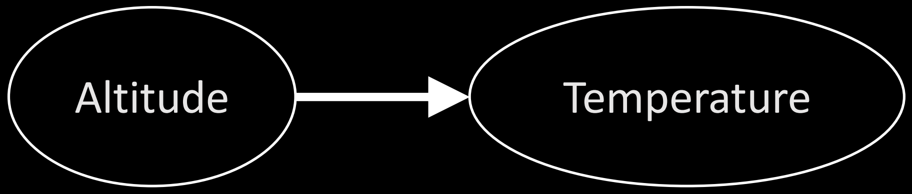
→ Definition

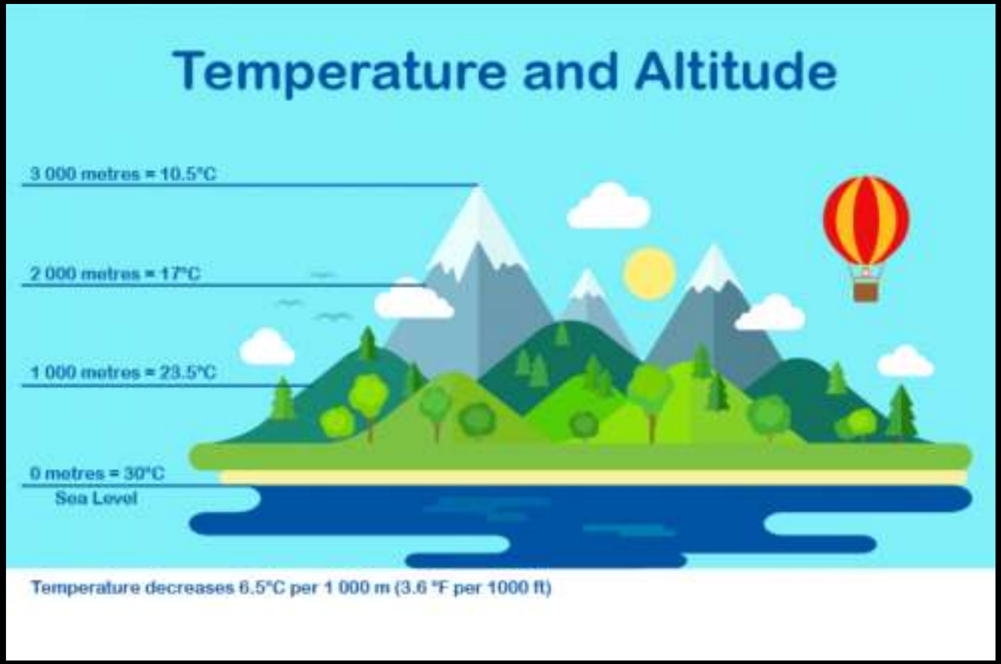
→ Experiments + Results

Conclusions

Future Work

Introduction to Causal Discovery





- Numerous relationships are causal
- How to learn such relationships?
- Can we build Causal ML?

Image-Source: https://letstalkscience.ca/educational-resources/backgrounders/weather-temperature

How to learn causal relationships?

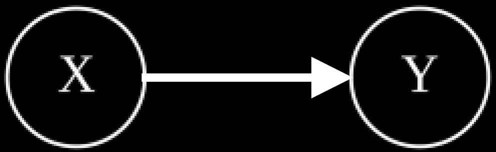
- Golden standard – randomized control trials or A/B tests
  - Can be expensive or infeasible
- Intervention or manipulation
  - Same problems
- How to learn causal relationship from observational data?
  - Closely related to structure learning in Bayes networks
- Multiple variables (Multivariate case)
  - Interaction between variables usually helps to learn the structure
- 2 variables (Bivariate case)
  - More difficult, additional assumptions are required

How to learn causal relationships?

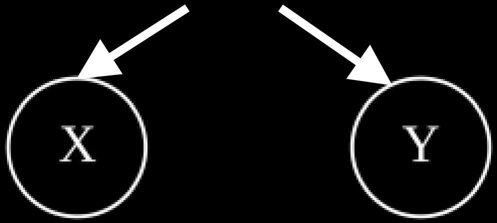
Judea, P. 2000. Causality: models, reasoning, and inference. Cambridge University Press.

If statistical association is observed, then one of the following holds

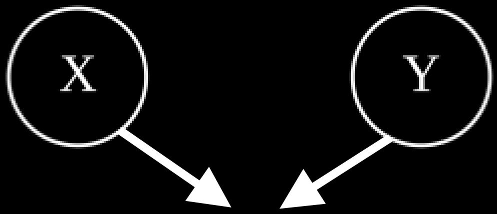
Causal relation



Confounder

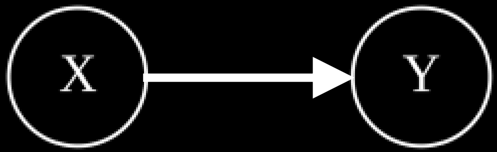


Selection bias



How to identify direction?

Structure learning in bivariate case: idea



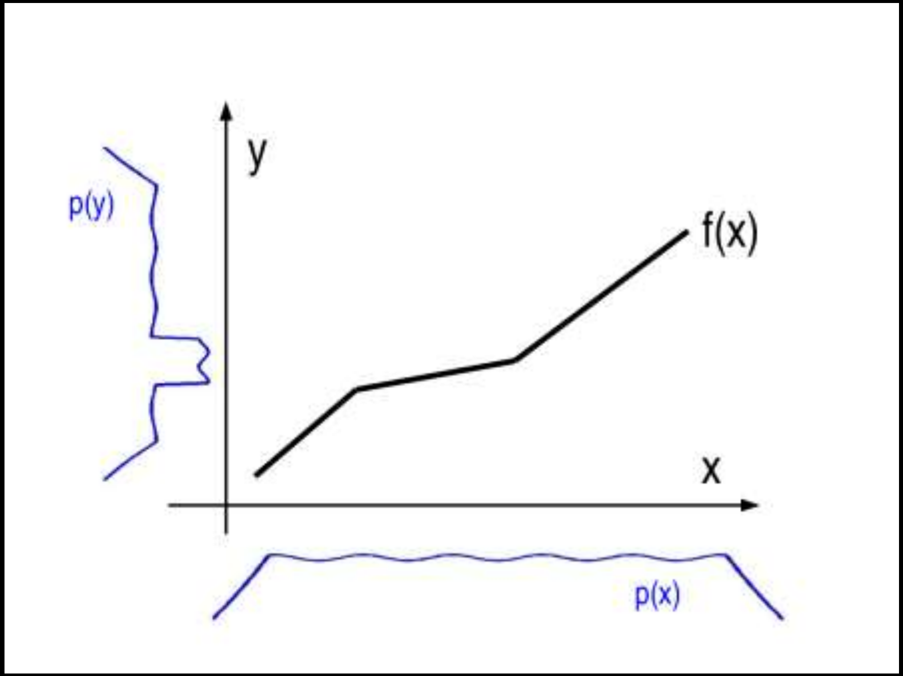
$$P(X,Y) \quad = \quad P(X) \cdot P(Y|X) \quad = \quad P(Y) \cdot P(X|Y)$$

$$P(X) \cdot P(Y|X) \quad < \quad P(Y) \cdot P(X|Y)$$

Model of lower complexity:

- Y contains information about X but not vice versa
- X has one source of randomness, Y has 2 sources of randomness

Structure learning in bivariate case: idea



Source: Janzing et al. (2012): “Information-geometric approach to inferring causal directions”

Additive noise models (ANM)

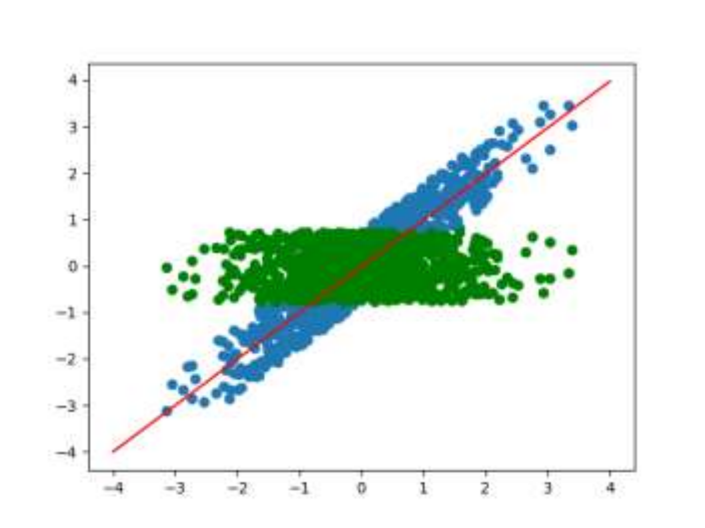
$$Y = X + \text{Noise}$$

Noise and X are independent

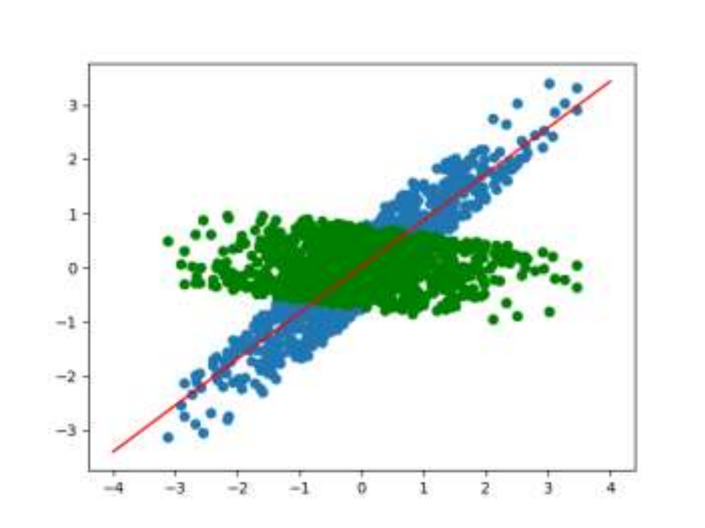
RESIT: Regression with Subsequent Independence Test
- No assumptions on the distribution type

RESIT

$Y := X + N_y$



$X := Y + N_x$



$C_{X \to Y} = \mathrm{Ind}(X, Y_{res})$

$C_{Y \to X} = \mathrm{Ind}(Y, X_{res})$

$$C_{X \to Y} > C_{Y \to X}$$

Kap B., Aleksandrova M., Engel T. - JFRB 2021

Independence estimators:

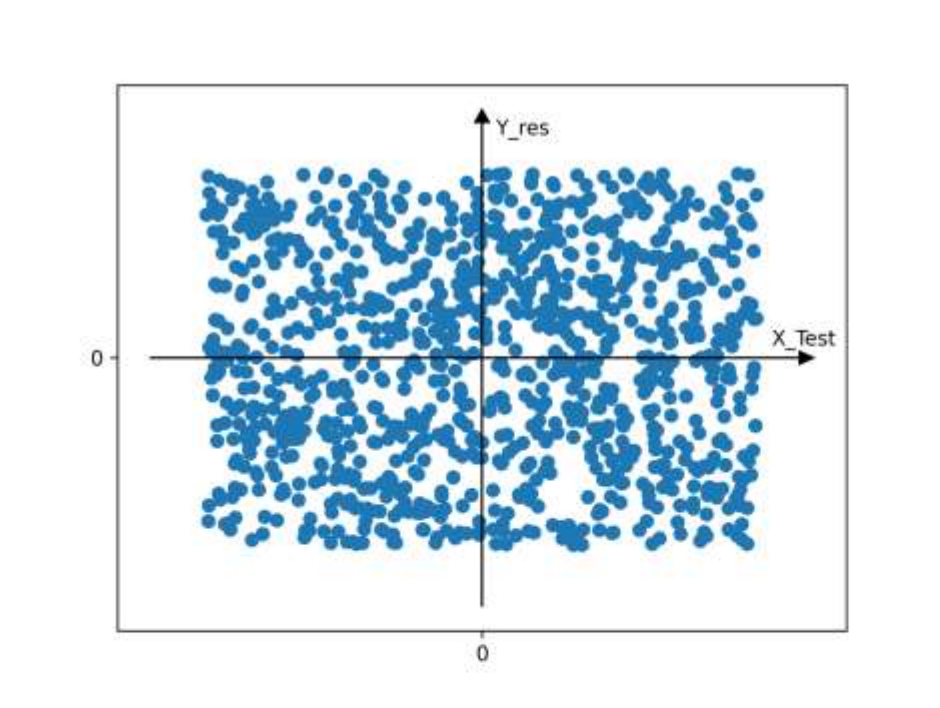
- Hilbert-Schmidt Independence Criterion (HSIC) with RBF Kernel
- HSIC using incomplete Cholesky decomposition with high precision
- HSIC using incomplete Cholesky decomposition with low precision
- Distance covariance
- Distance correlation
- Hoeffding’s Phi

Entropy estimators:

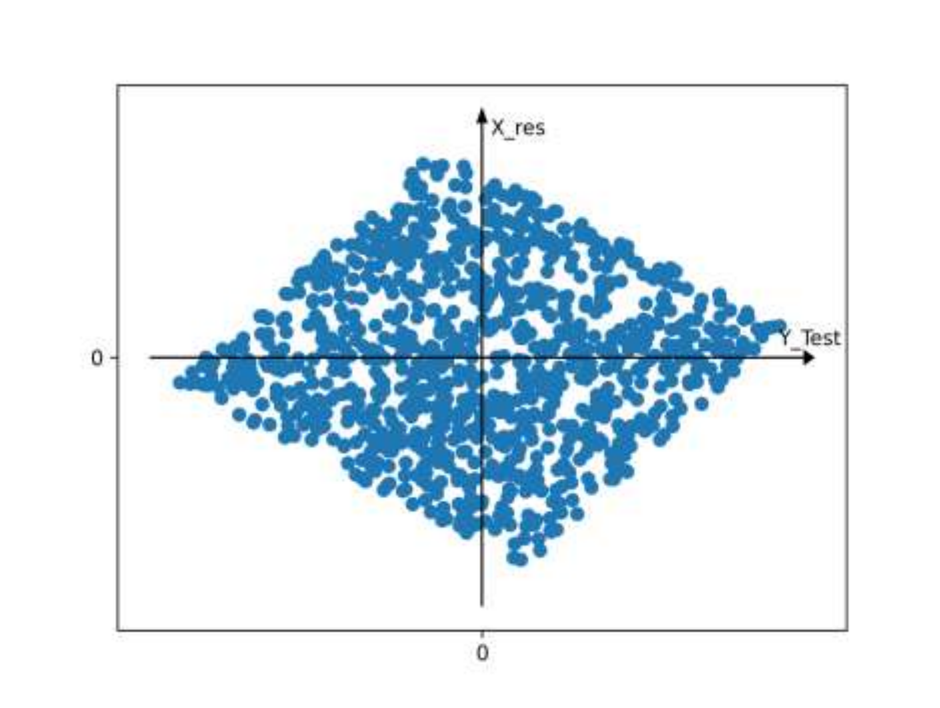
- Shannon differential entropy using k-nearest neighbours with k=3
- Shannon differential entropy using k-nearest neighbours with k=3 and kd-tree
- Shannon differential entropy using k-nearest neighbours with k=3
- Maximum entropy distribution based Shannon entropy estimator
- Maximum entropy distribution based Shannon entropy estimator, different parameters
- Shannon entropy estimator using Vasicek’s spacing method

RESIT

$Y := X + N_y$



$X := Y + N_x$

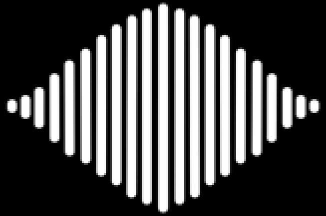
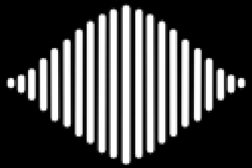
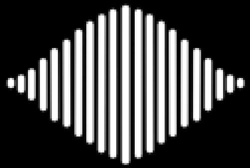


What is the impact of noise level?

Kap B., Aleksandrova M., Engel T. - JFRB 2021

Experimental setup

$$i \in \{0.01, 0.02, \dots, 1.00\} \cup \{1, 2, \dots, 100\}$$

 =  + 

$$Y = X + \text{Noise}$$

Kap B., Aleksandrova M., Engel T. - JFRB 2021

Experimental setup

$Y = X + \text{Noise} \;\; (X \rightarrow Y)$

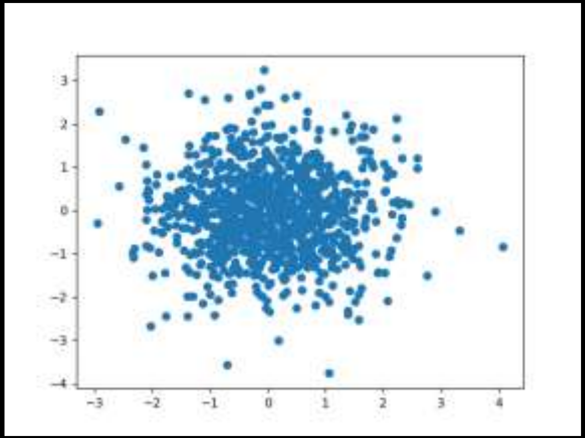
$Y = X^3 + \text{Noise}$

$X \sim N(0, 1)$ or $U(-1, 1)$ or $L(0, 1)$

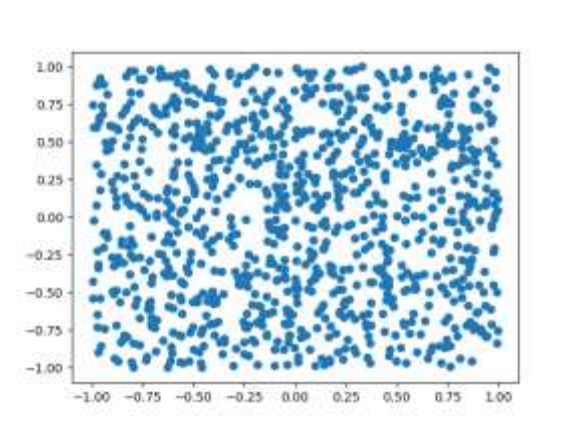
Noise ~ N, $1 \cdot i$) or $U(-1 \cdot i, 1 \cdot i)$ or $L(0, 1 \cdot i)$

18 Combinations in total

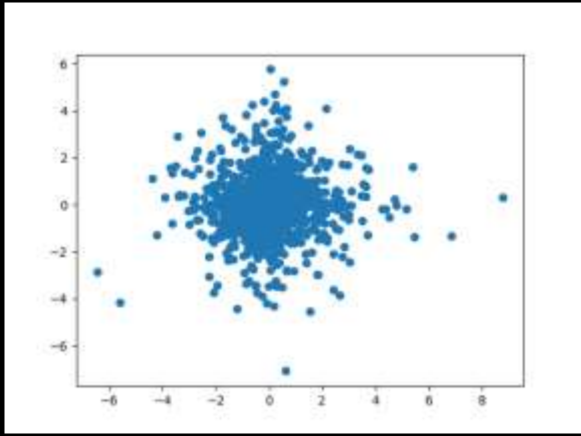
Gaussian



Uniform



Laplace



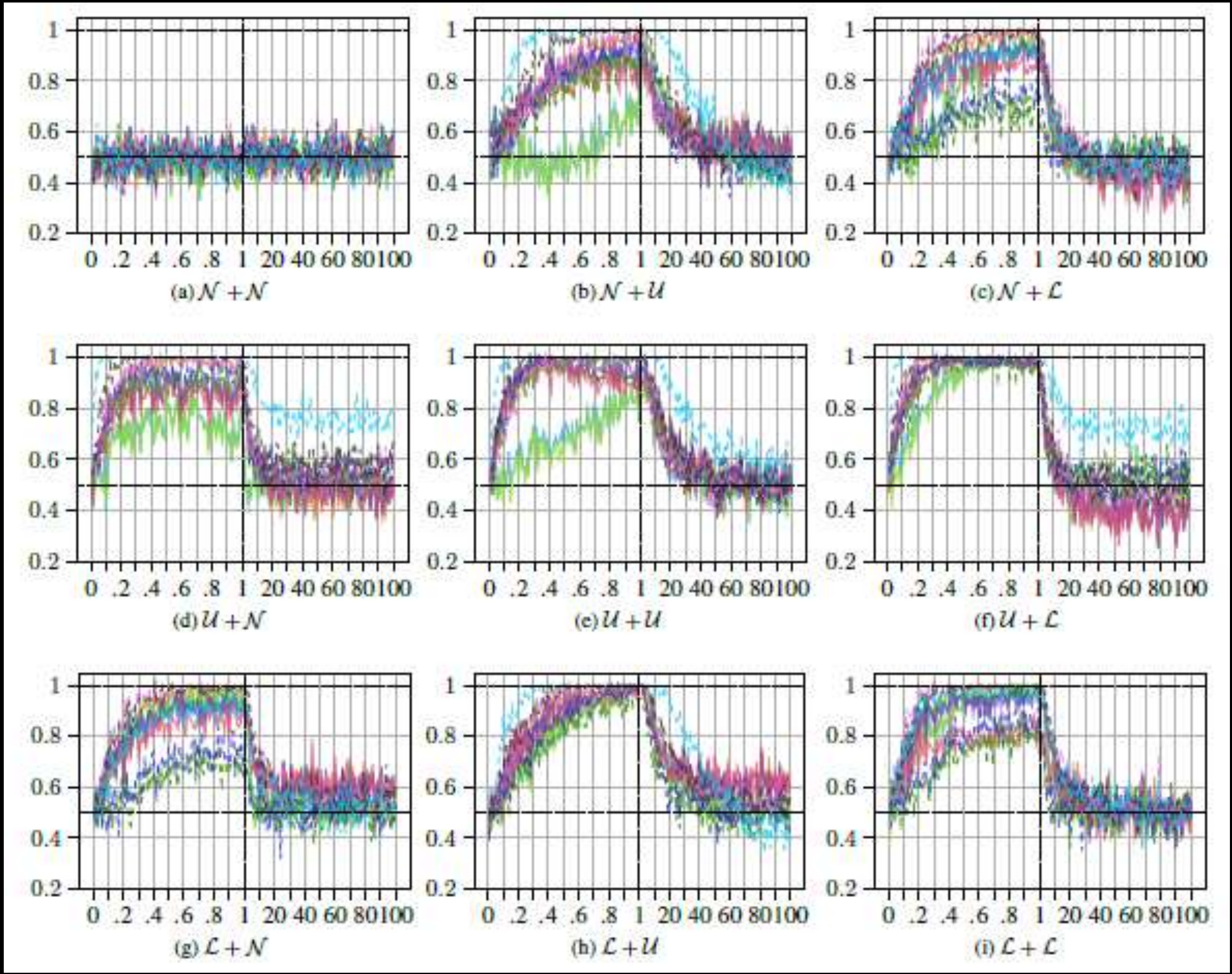
100 tests / 1000 new samples for each test

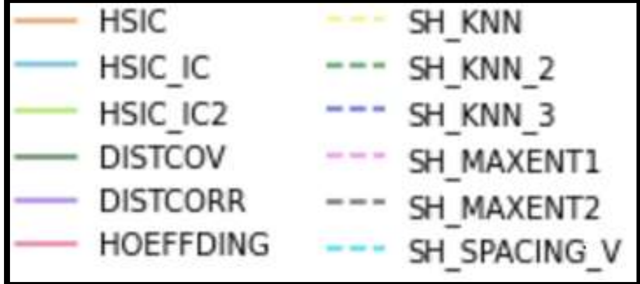
Evaluation metric – accuracy

Kap B., Aleksandrova M., Engel T. - JFRB 2021

Y = X + Noise

Accuracy as a function of *i*

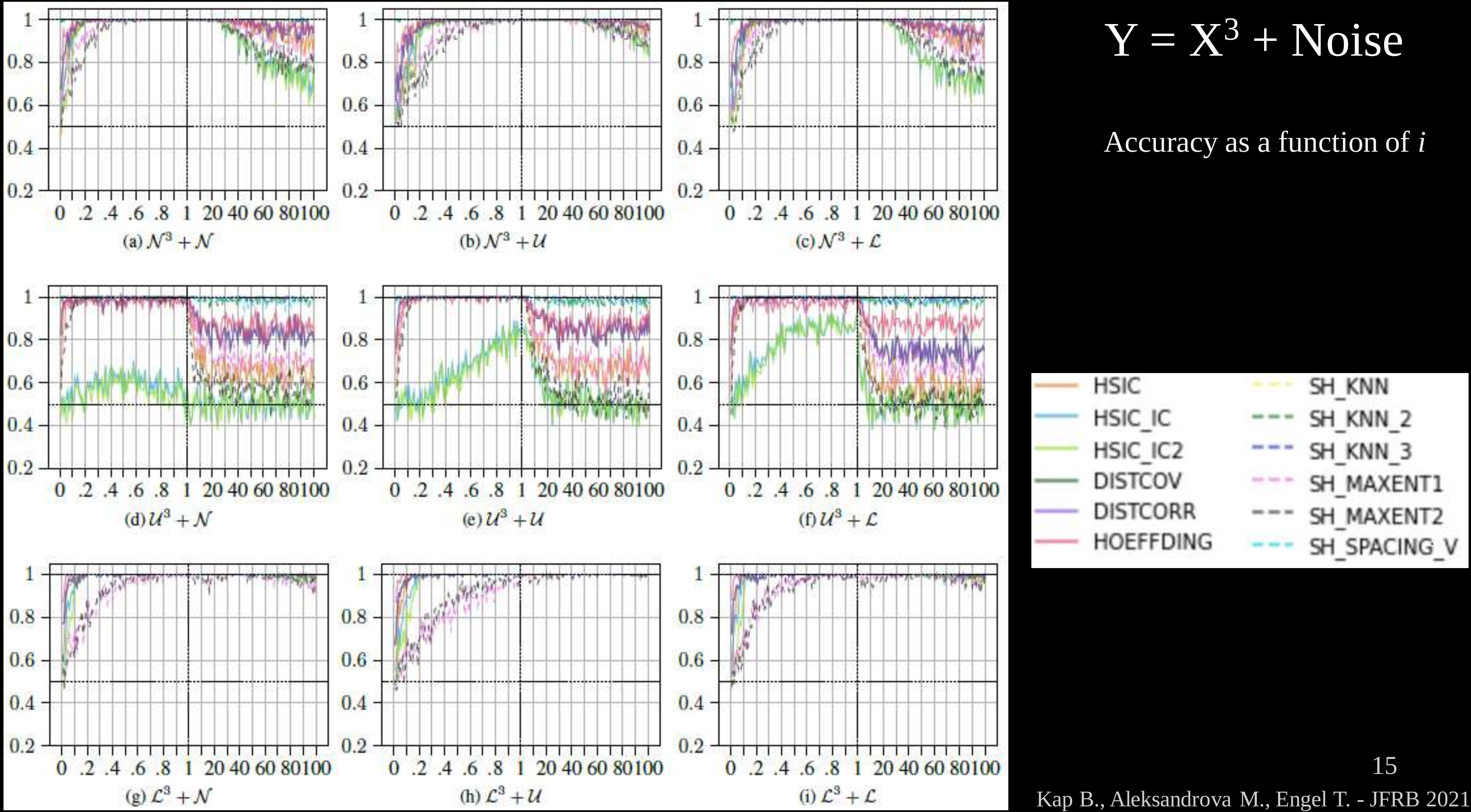




Kap B., Aleksandrova M., Engel T. - JFRB 2021

$Y = X^3 + \text{Noise}$

Accuracy as a function of *i*



Kap B., Aleksandrova M., Engel T. - JFRB 2021

Conclusions

- Different noise levels → Impact identifiability
  - Significantly small or big → unidentifiable models
  - "Significantly" is different for different setups
- Recommendations:
  - Best and worst independence estimators:
    - HSIC with RBF Kernel
    - HSIC with Cholesky Decomposition
  - Best and worst entropy estimators:
    - Shannon E. with Vasicek's spacing method
    - Maximum entropy dist. based Shannon entropy estimator
  - Model-specific recommendations

Future work

- Theoretical analysis of estimators → which recommendations can be made?
- Analytical formalization of noise impact
- Generalization of the results for other types of distributions and their combinations