Joint International Scientific Conferences on AI
BNAIC/BENELEARN 10-12.10.2021

The Effect of Noise Level on the Accuracy of Causal Discovery Methods with Additive Noise Models

Benjamin KAP, Dr. Marharyta ALEKSANDROVA, Prof. Thomas ENGEL

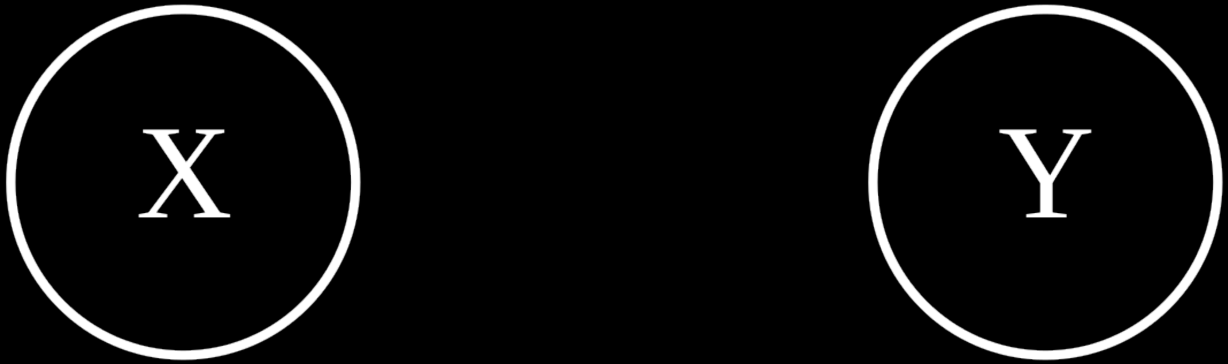
SECAN-Lab

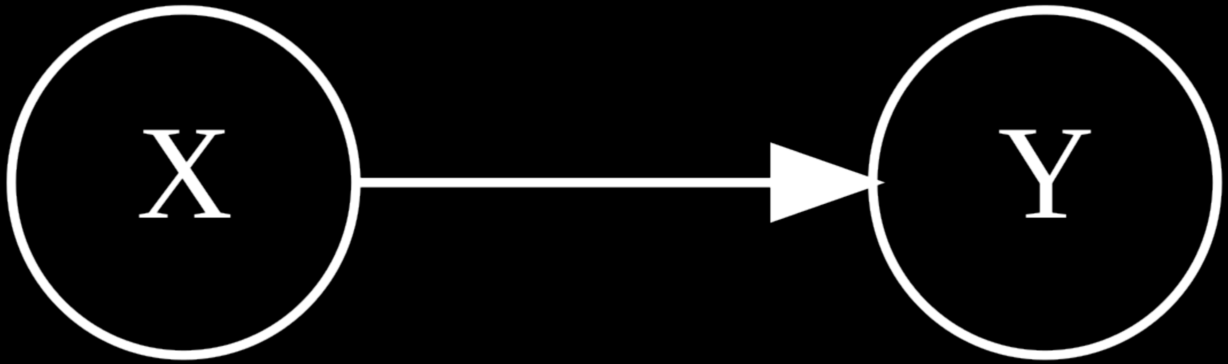
uni.lu UNIVERSITÉ DU LUXEMBOURG

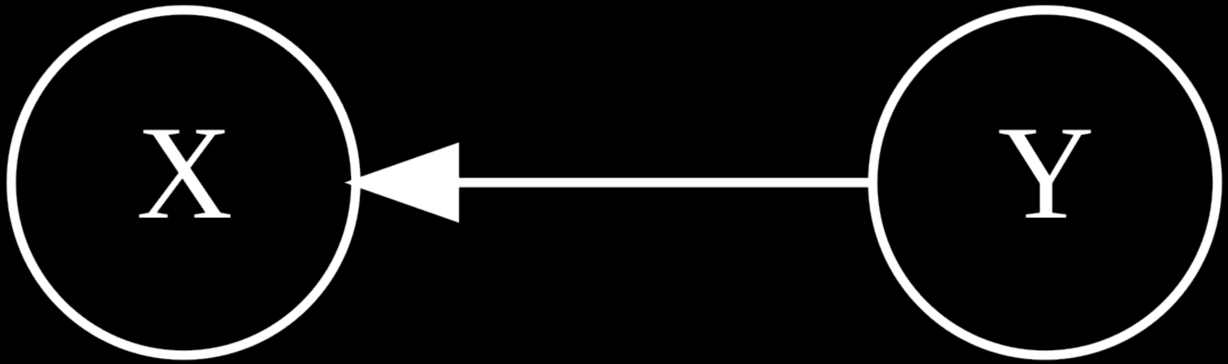
Outline

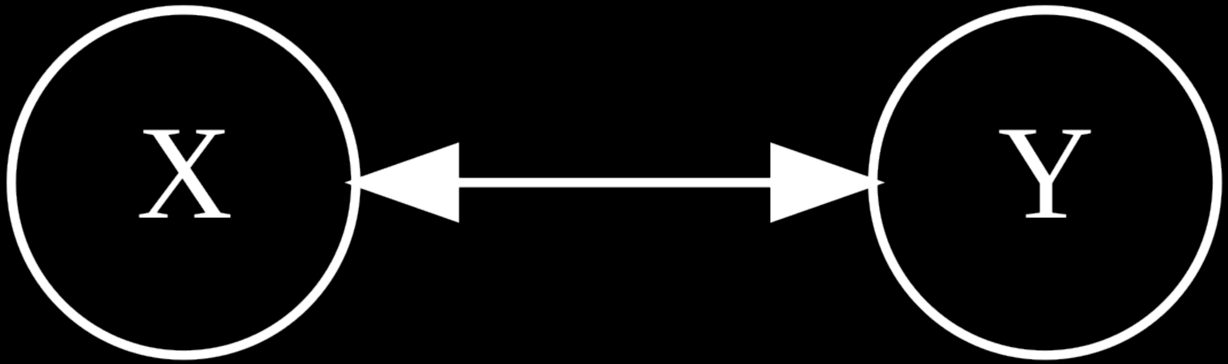
Introduction to Causal Discovery

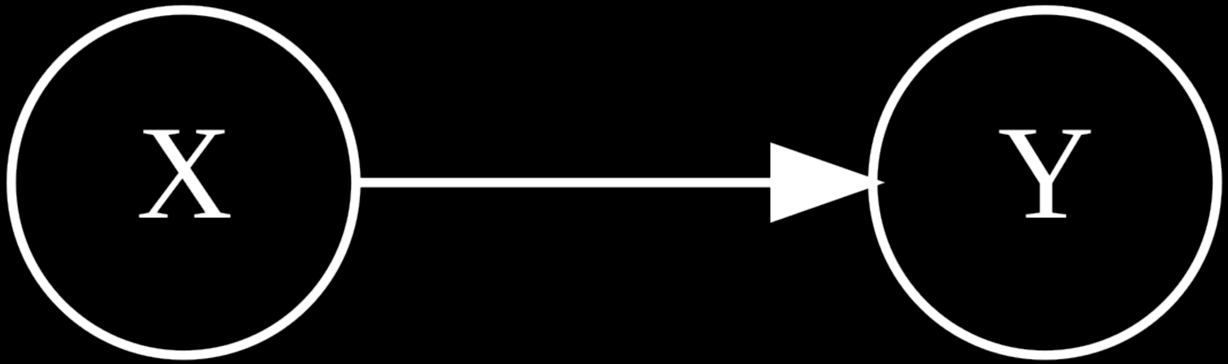
What is Causal Discovery?


X
Y


X
Y

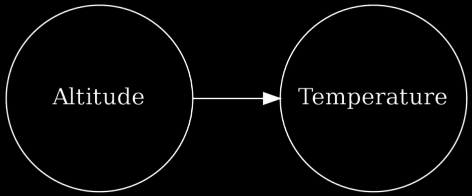

X
Y


X
Y



Cause

Effect



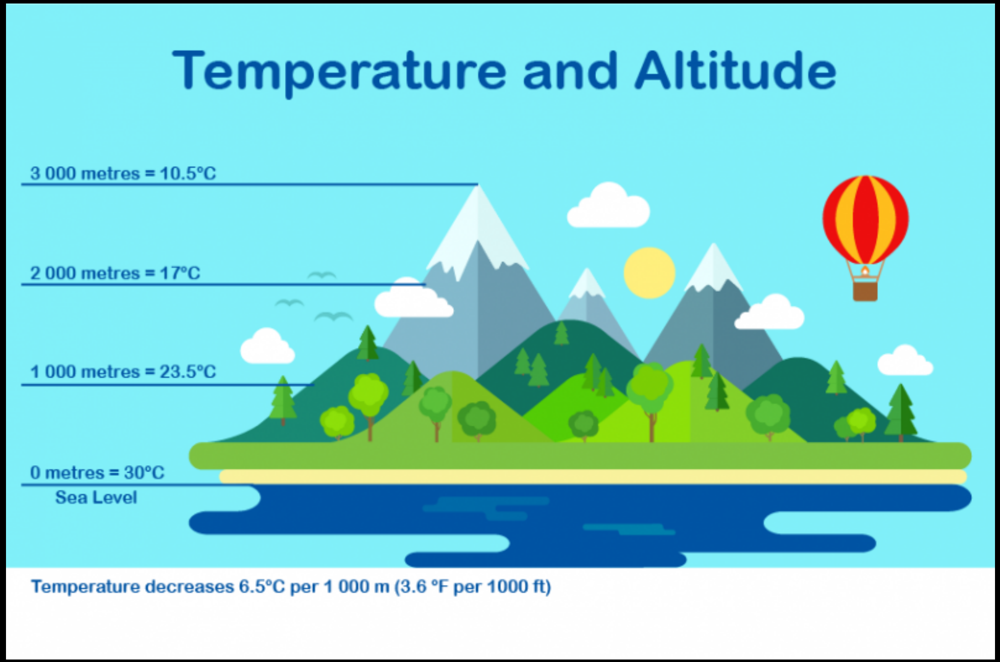
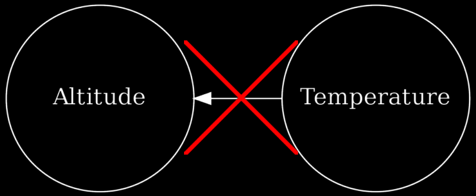


Image-Source: https://letstalkscience.ca/educational-resources/backgrounders/weather-temperature



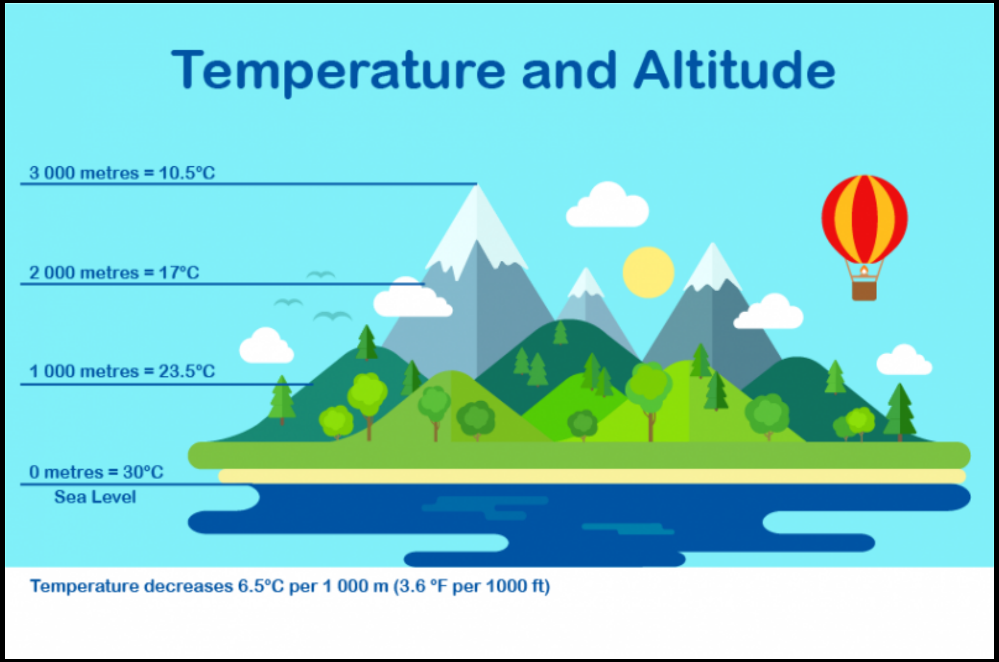


Image-Source: https://letstalkscience.ca/educational-resources/backgrounders/weather-temperature

Importance

Controlled Tests $\rightarrow$ A/B Tests

Observational Data only

State of the Art

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

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

$$P(X) \cdot P(Y|X) > P(Y) \cdot P(X|Y)$$

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

Very complex

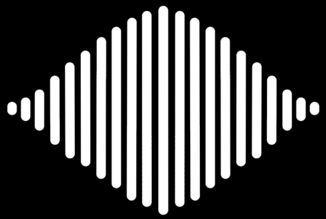
A lot of assumptions

Additive Noise Models (ANM)

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

Not enough attention to noise levels in the noise term!

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

 =  + 

RESIT

&

Uncertainty Scoring

RESIT

Regression with Subsequent Independence Test

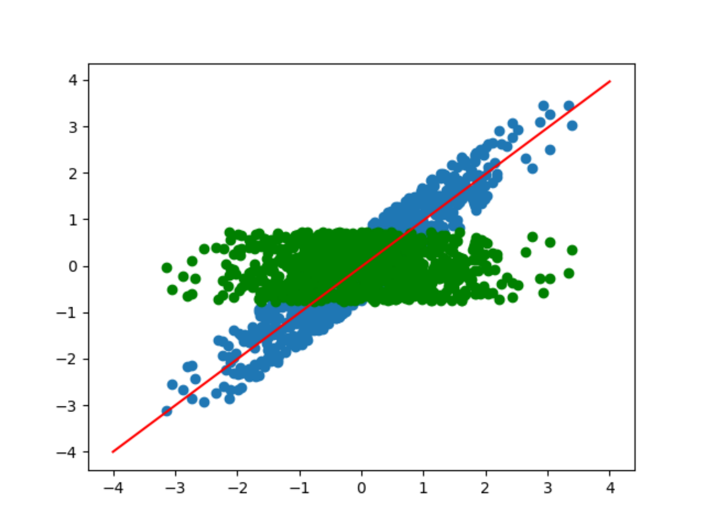
No assumption on the distribution type

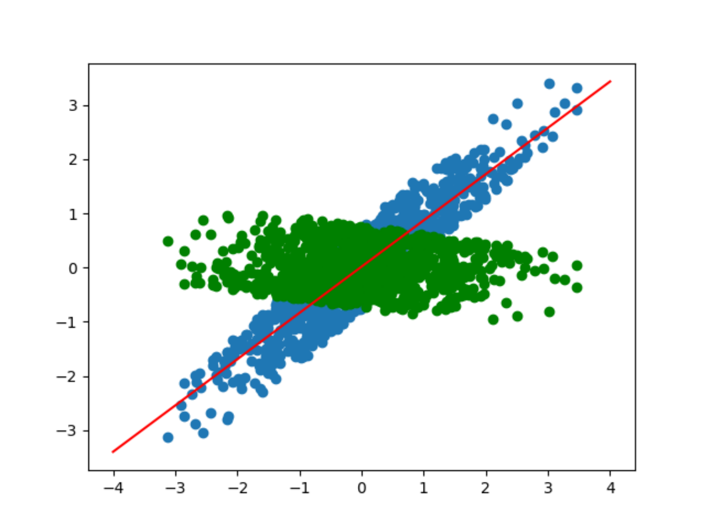
Optimization problem is generally non-convex

$Y := X + N_y$

RESIT

$X := Y + N_x$





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

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

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

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

Assumption: $(X \rightarrow Y)$ xor $(Y \rightarrow X)$

$$\rightarrow C_{X\rightarrow Y} \qquad ? \qquad C_{Y\rightarrow X}$$

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

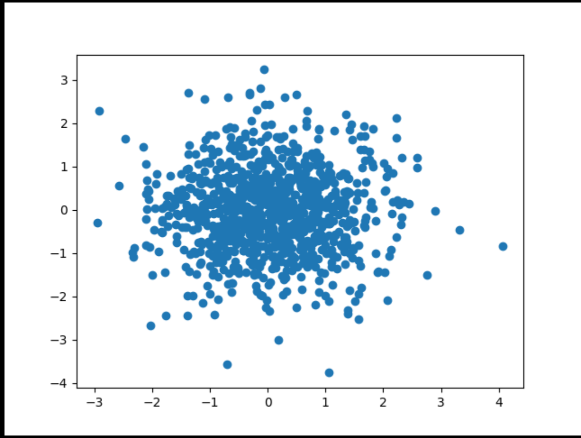
Entropy estimators:

- 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

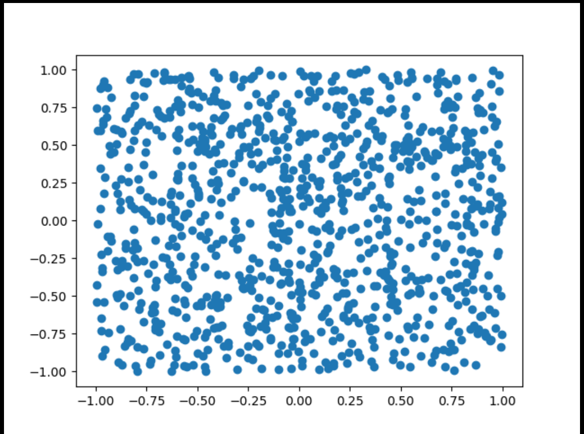
$$Y = X + \text{Noise} \quad (X \rightarrow Y)$$

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

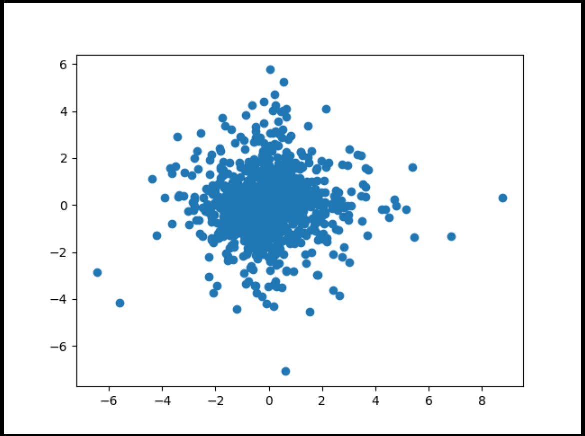
Gaussian



Uniform



Laplace



X ~ Gaussian(0, 1) or Uniform(-1, 1) or Laplace(0, 1)

Noise ~ Gaussian(0, $1 \cdot i$) or Uniform($-1 \cdot i$, $1 \cdot i$) or Laplace(0, $1 \cdot i$)

"i-factor"

18 Combinations in total

RESIT

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

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

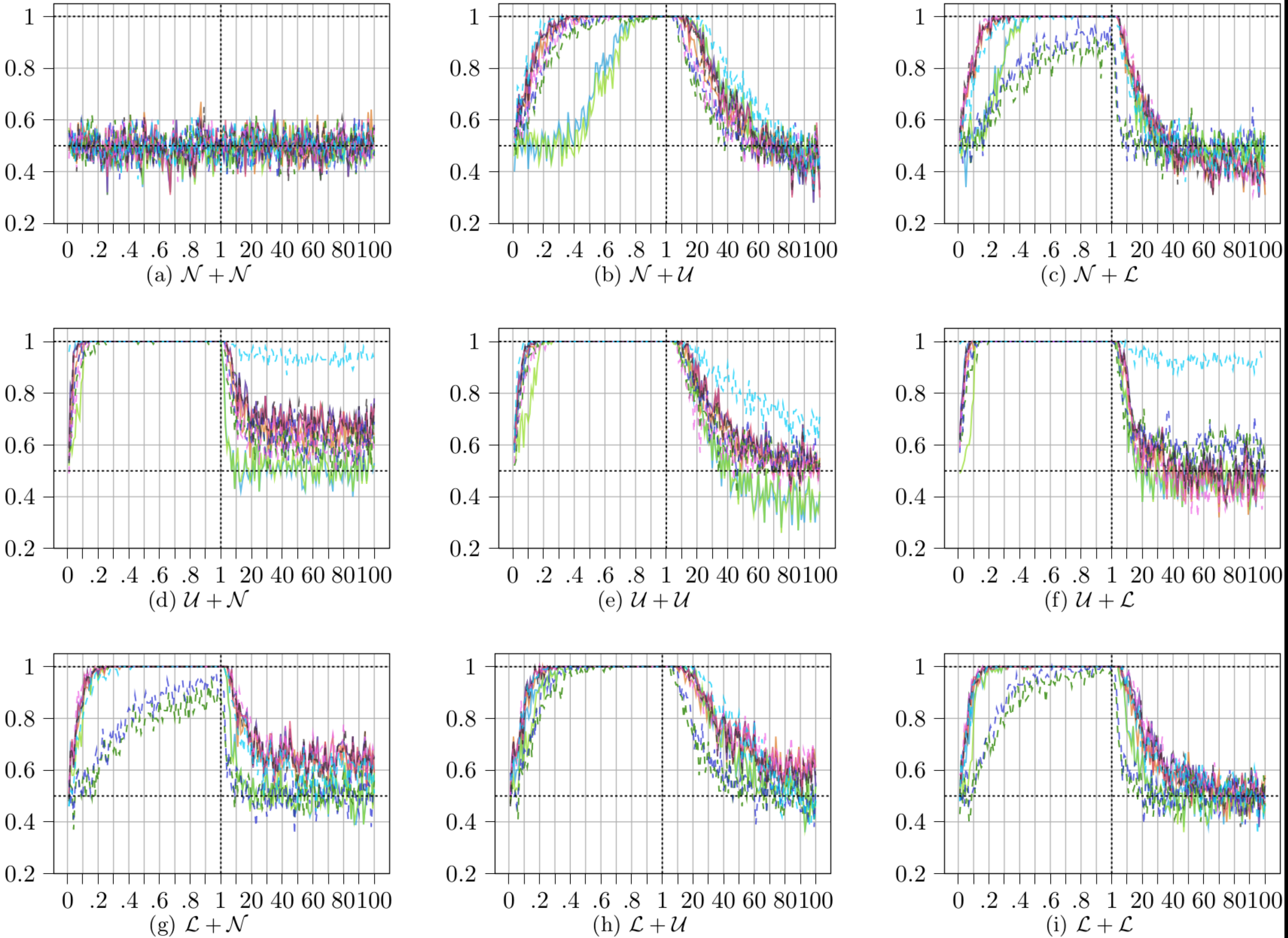
Example: Y = Gaussian(0, 1) + Laplace(0, 20)

Test RESIT 100 times

1000 new samples for each test

Y = X + Noise

Accuracy as a function of i

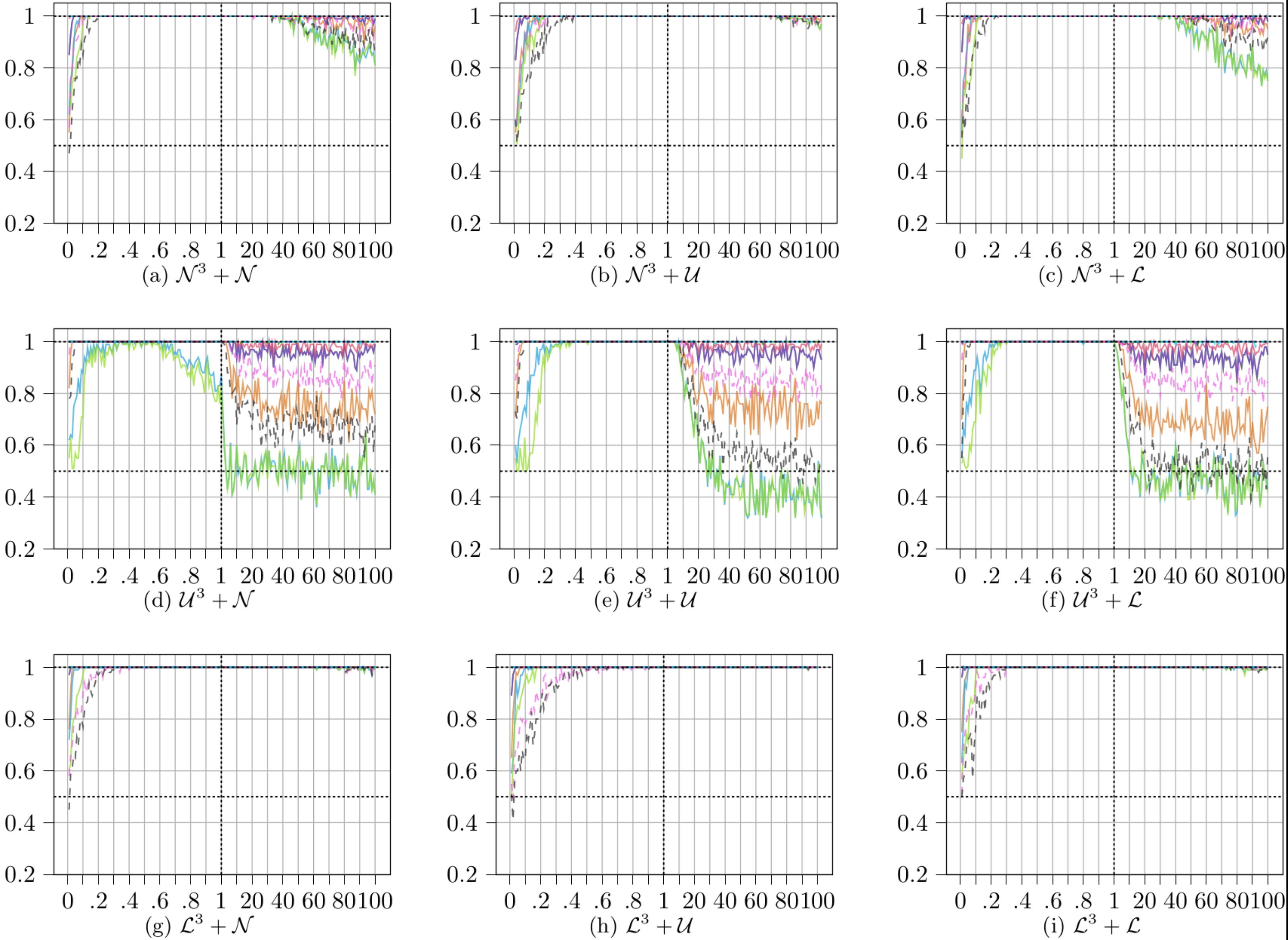


Kap B.,
Aleksandrova M.,
Engel T.

BNAIC/BENELEARN 2021

Y = X³ + Noise

Accuracy as a function of i



Kap B., Aleksandrova M., Engel T.
BNAIC/BENELEARN 2021

Uncertainty Scoring

No assumption on the distribution type

Y = Gaussian + Gaussian

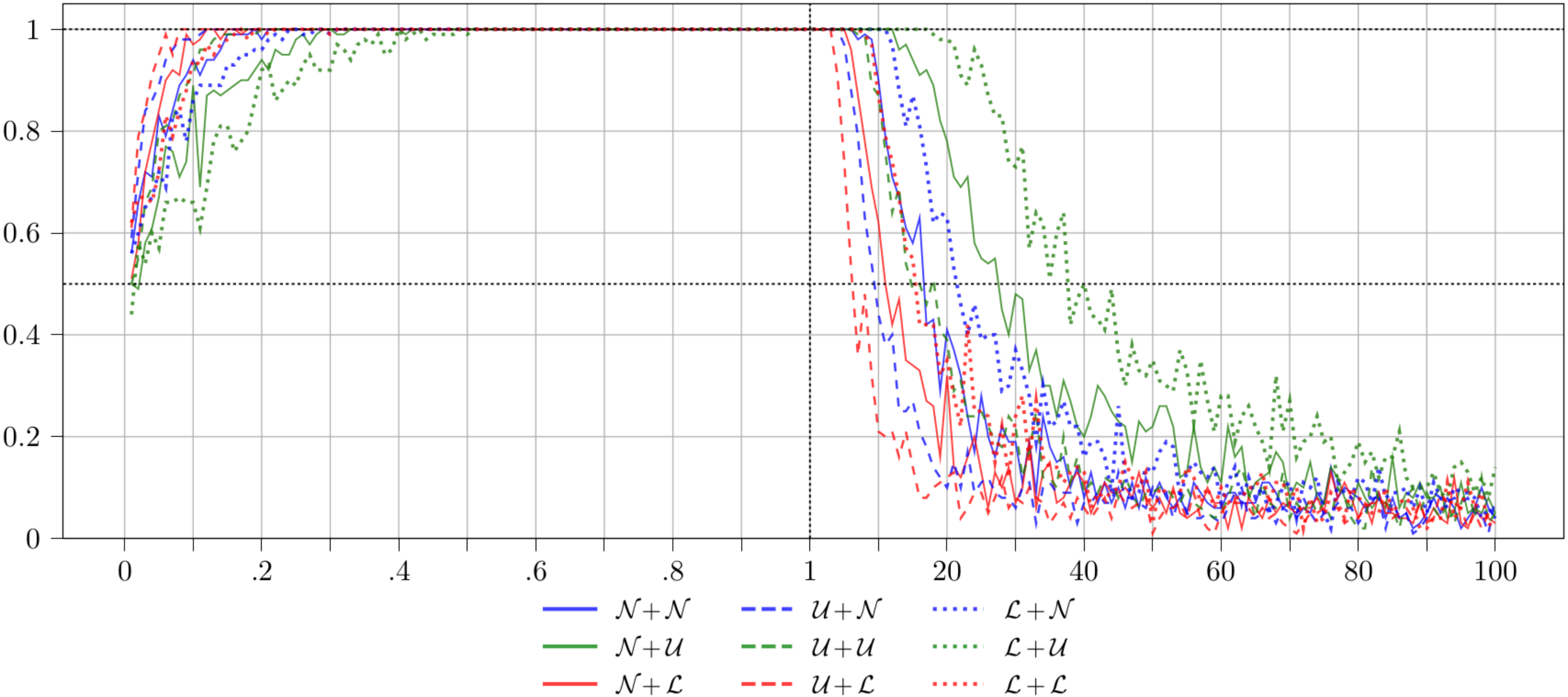
Uncertainty Scoring

Y = X + Noise

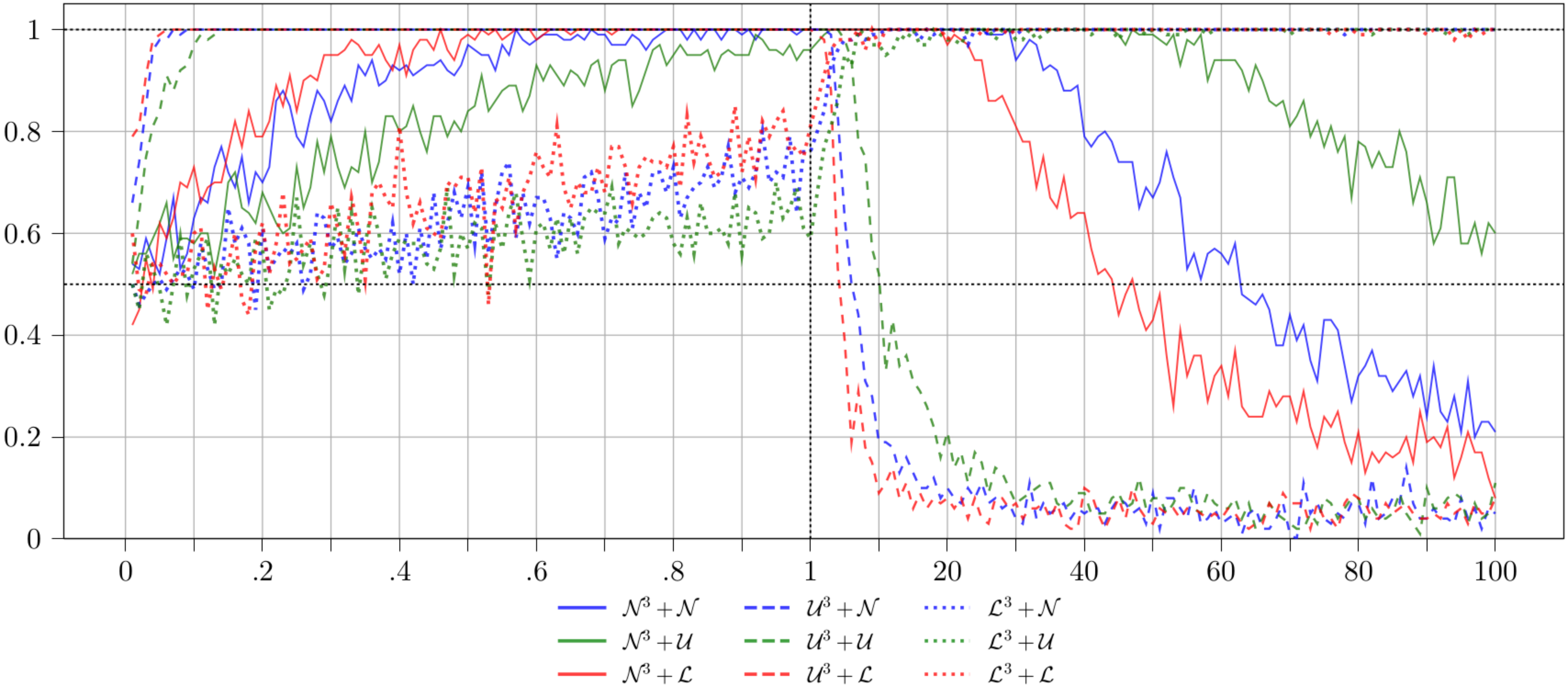
Conditional Fisher's Independence Test

18 Combinations (Linear + Non-Linear + Distribution types)

100 tests; 1000 new samples



Y = X + Noise; Accuracy as a function of i



Y = X^3 + Noise; Accuracy as a function of i

Conclusion

Different noise levels $\rightarrow$ Impact!

Different distribution types

Significantly small or big $\rightarrow$ Unidentifiability

Future Work

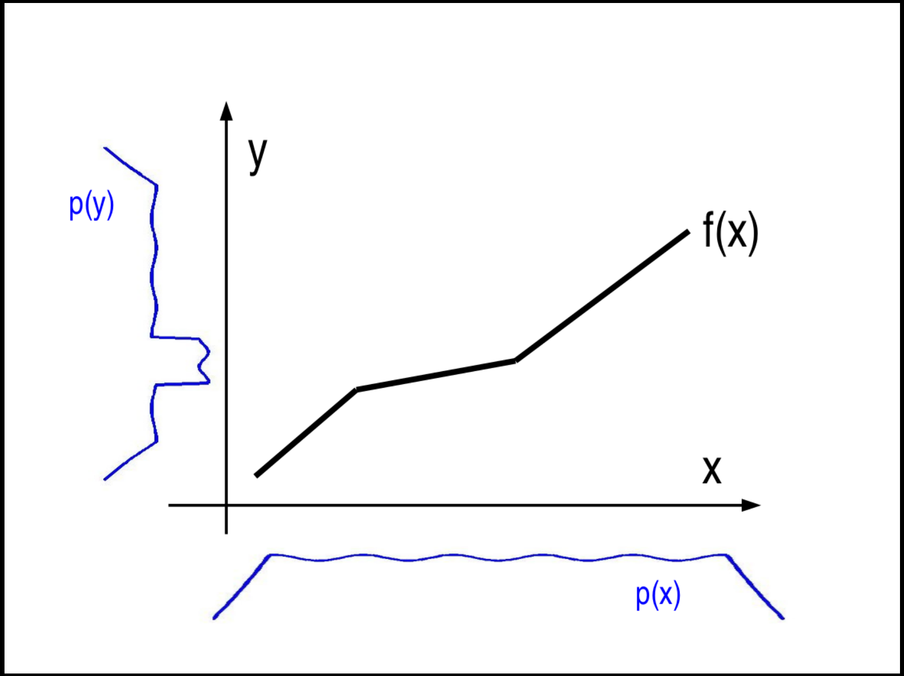
Point of failure $\rightarrow$ Estimators

Formalization

Generalization

Thank you!

State of the Art



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

Best and Worst Ind. Estimators:

HSIC with RBF Kernel
HSIC with Cholesky Decomposition

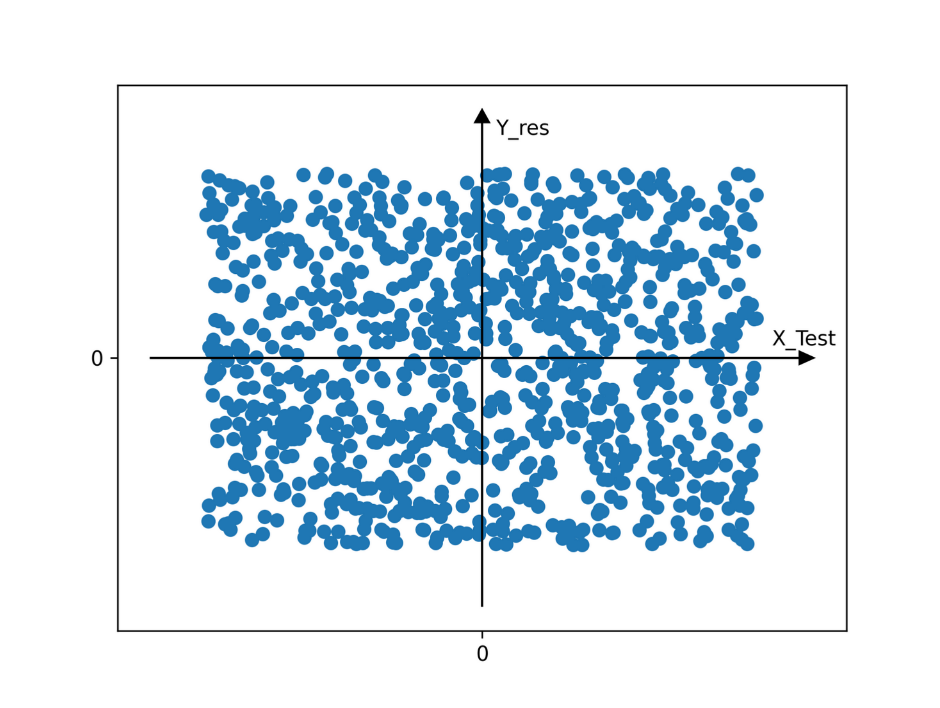
Best and Worst Entropy Estimators:

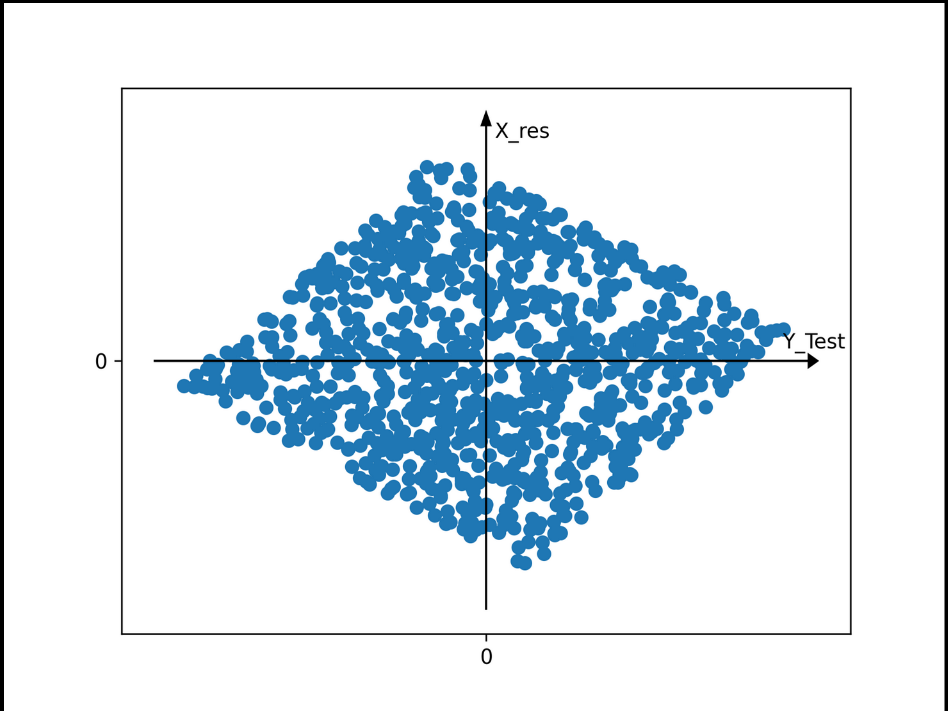
Shannon E. with Vasicek's spacing method
Maximum entropy dist. based Shannon entropy estimator

$$Y := X + N_y$$

RESIT

$$X := Y + N_x$$





$Y = \beta \cdot X + \text{Noise}$ $\quad$ $\text{Var}(Y)/\text{Var}(X) > 1 - \beta^2$

1) $\text{Var}(Y) = E(\text{Var}(Y|X)) + \text{Var}(E(Y|X)) = \text{Var}(Y) + \beta^2\text{Var}(X) > \text{Var}(X)$

2) $E(\text{Var}(X|Y)) = \text{Var}(X) – \text{Var}(E(X|Y))$

$= \text{Var}(X) - (\beta^2\text{Var}(X)^2) / (\beta^2\text{Var}(X) + \text{Var}(Y)) < \text{Var}(Y)$

$= E(\text{Var}(Y|X))$

$$\text{Set } S = \{ X, Y, W, V \}$$

$$\text{Ordering } \pi = [V, X, W, Y]$$

Ordering $\pi = [\ V,\ X, W, Y\ ]$

Ind(W,V): Independent

Ind(W,X): Dependent

→ X is a parent of W!