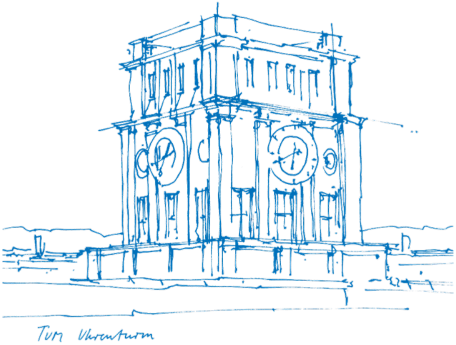
Munich Center for Machine Learning
TUM School of Computation, Information and Technology
Technical University of Munich

Munich Center for Machine Learning

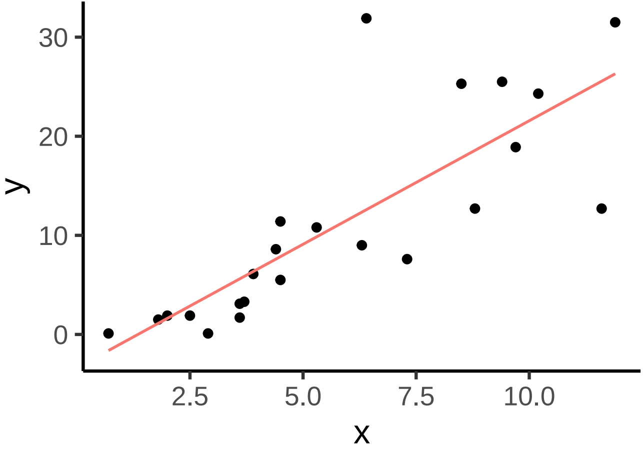
Confidence in Causal Inference under Structure Uncertainty

David Strieder (joint work with Mathias Drton)

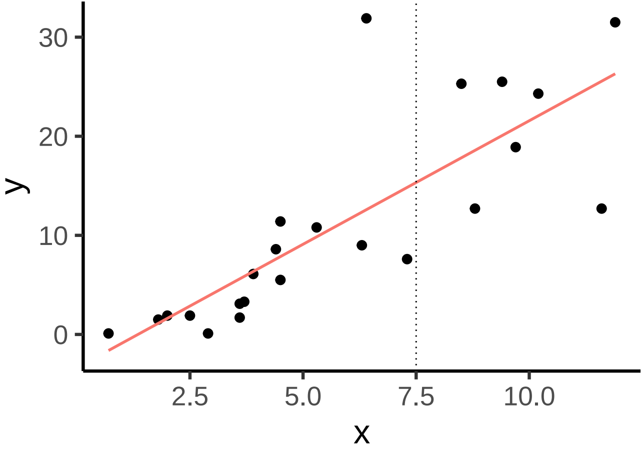
18. Doktorand:innentreffen der Stochastik, Heidelberg
Aug 21 to Aug 23, 2023



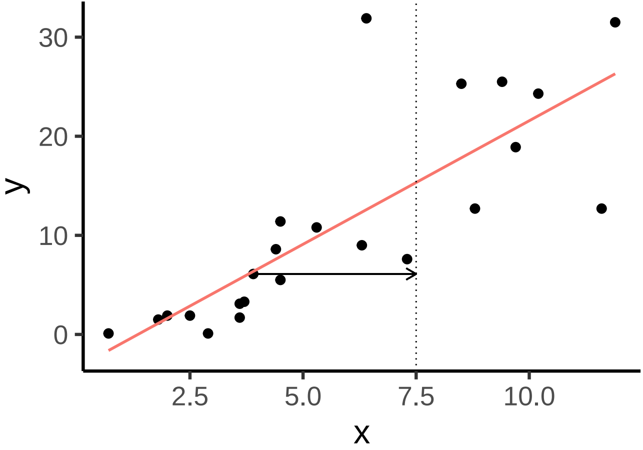
Motivation



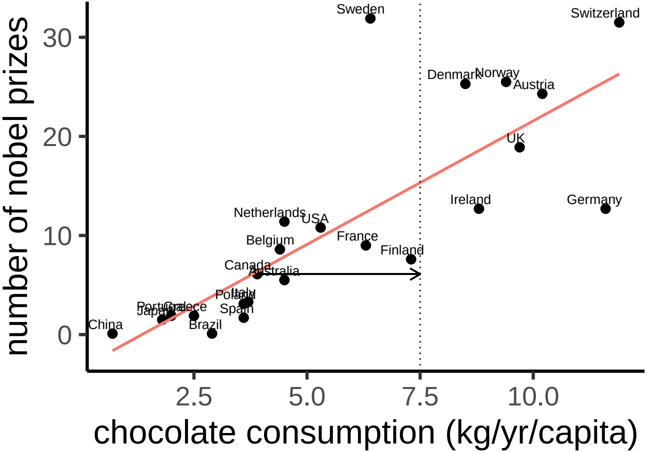
Motivation



Motivation



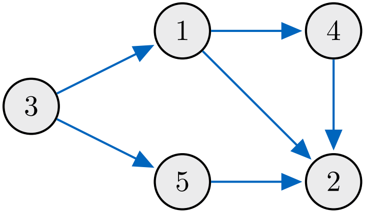
Motivation



Data from Messerli F., *Chocolate Consumption, Cognitive Function, and Nobel Laureates*, N Engl J Med 2012.

Structural Equation Models
Observational Distribution

$$
\begin{aligned}
X_1 &= f_1(X_3, \varepsilon_1) \\
X_2 &= f_2(X_1, X_4, X_5, \varepsilon_2) \\
X_3 &= f_3(\varepsilon_3) \\
X_4 &= f_4(X_1, \varepsilon_4) \\
X_5 &= f_5(X_3, \varepsilon_5)
\end{aligned}
$$

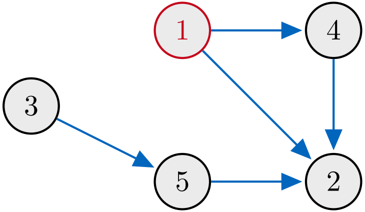


where ε_j mutually independent

Structural Equation Models

Interventional Distribution $do(X_1 = x_1)$

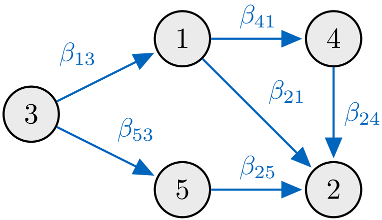
$$
\begin{aligned}
X_1 &= x_1 \\
X_2 &= f_2(X_1, X_4, X_5, \varepsilon_2) \\
X_3 &= f_3(\varepsilon_3) \\
X_4 &= f_4(X_1, \varepsilon_4) \\
X_5 &= f_5(X_3, \varepsilon_5)
\end{aligned}
$$



where ε_j mutually independent

Structural Equation Models
Linear SEM with equal variances

$$X_1 = \beta_{13} X_3 + \varepsilon_1$$
$$X_2 = \beta_{21} X_1 + \beta_{24} X_4 + \beta_{25} X_5 + \varepsilon_2$$
$$X_3 = \varepsilon_3$$
$$X_4 = \beta_{41} X_1 + \varepsilon_4$$
$$X_5 = \beta_{25} X_3 + \varepsilon_5$$

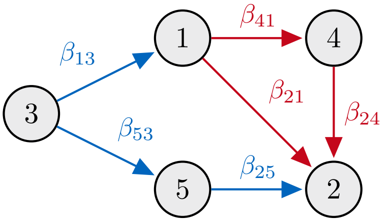


where $\varepsilon_j \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2)$

Setup

- **Interest:** Total causal effect of an intervention on X_1 onto X_2.

$$\mathcal{C}(1 \to 2) := \frac{\mathrm{d}}{\mathrm{d}x_1}\mathbb{E}[X_2 | \mathrm{do}(X_1 = x_1)]$$
$$= \Sigma_{12|p(1)}/\Sigma_{11|p(1)}$$



$$\mathcal{C}(1 \to 2) = \beta_{21} + \beta_{41}\beta_{24}$$

Setup

- **Idea:** Use test inversion.

Tests for $\mathsf{H}_0^{(\psi)} : \mathcal{C}(1 \to 2) = \psi$ → Confidence interval for $\mathcal{C}(1 \to 2)$

- **Goal:** Construct suitable **tests for all** $\psi \in \mathbb{R}$.

Setup

- **Idea:** Use test inversion.

Tests for $\mathsf{H}_0^{(\psi)} : \mathcal{C}(1 \to 2) = \psi$ → Confidence interval for $\mathcal{C}(1 \to 2)$

- **Goal:** Construct suitable **tests for all** $\psi \in \mathbb{R}$**.**

- **Difficulty:** Hypothesis is **union of single hypotheses** over all structures.

$$\mathsf{H}_0^{(\psi)} := \bigcup_{G \in \mathcal{G}(d)} \mathsf{H}_0^{(\psi)}(G)$$

Single Hypothesis $\mathsf{H}_0^{(\psi)}(G)$

$$\mathsf{H}_0^{(\psi)}(G): \left\{\Sigma \in \mathsf{PD}(d) : \exists \sigma^2 > 0 \text{ with } \psi\sigma^2 = \Sigma_{1,2|p(1)} \text{ and } \sigma^2 = \Sigma_{k,k|p(k)} \ \forall\, k = 1, \ldots, d\right\}$$

- **Idea:** Use theory of intersection union test.
- Reject union if we reject each single hypothesis.

Constrained likelihood ratio test

- **Idea:** Relax alternative to entire cone of covariance matrices.
- Each single defines hypothesis smooth submanifold.
- Limit distribution is a chi-squared distribution.
- **Result:** Asymptotic $(1-\alpha)$-confidence set for causal effect $\mathcal{C}(1 \to 2)$ is

$$\{\psi \in \mathbb{R} : \min_{G \in \mathcal{G}(d)\,:\,1 <_G 2} \lambda_n^{(\psi)}(G) \leq \chi^2_{d,1-\alpha}\} \cup \{0 : \min_{G \in \mathcal{G}(d)\,:\,2 <_G 1} \lambda_n^{(0)}(G) \leq \chi^2_{d-1,1-\alpha}\}$$

Split likelihood ratio test[1]

- **Idea:** Split data and use universal critical value.
- Calculate MLE of Σ under alternative based on data part 1.
- Calculate MLE of Σ under hypothesis and likelihoods based on data part 2.
- **Result:** $(1-\alpha)$-confidence set for causal effect $\mathcal{C}(1 \to 2)$ is

$$\{\psi \in \mathbb{R} : \min_{G \in \mathcal{G}(d) : 1 <_G 2} \tilde{\lambda}_n^{(\psi)}(G) \leq -2\log(\alpha)\} \cup \{0 : \min_{G \in \mathcal{G}(d) : 2 <_G 1} \tilde{\lambda}_n^{(0)}(G) \leq -2\log(\alpha)\}$$

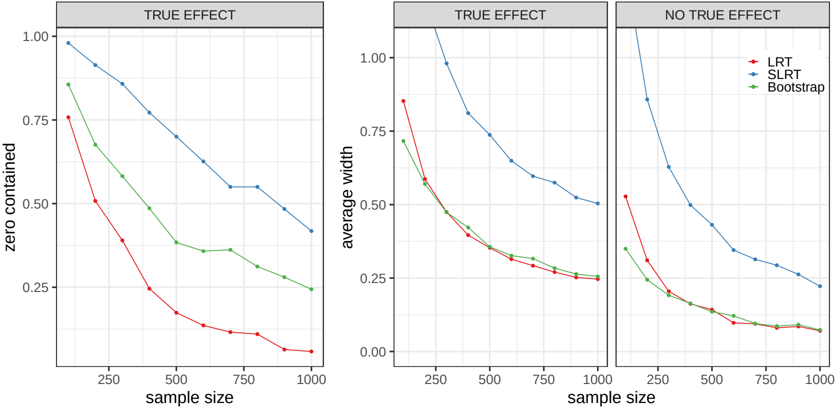
[1] Wasserman L, Ramdas A, Balakrishnan S., *Universal inference,* Proc. Natl. Acad. Sci. USA 2020.

Simulations

method	$n \backslash \beta$	TRUE EFFECT 0.05	0.1	0.5
LRT	100	1.00	0.99	1.00
	500	0.95	0.98	1.00
	1000	0.94	0.97	1.00
SLRT	100	1.00	1.00	1.00
	500	0.98	0.99	1.00
	1000	0.96	0.99	1.00
Bootstrap	100	0.64	0.71	0.97
	500	0.70	0.79	0.98
	1000	0.76	0.83	0.98

Empirical Coverage of 95%-CIs

Simulations



Zero Contained and Mean Width of 95%-CIs (1000 Random 5-dim. DAGs)