



# Causal Inference Theory and Applications in Enterprise Computing

Christopher Hagedorn, Johannes Huegle, Dr. Michael Perscheid

May 20, 2020

# Agenda

May 20, 2020

---

- **Recap: Causal Inference in a Nutshell**

- Causal Structure Learning

- **Jupyter Lab**

1. Causal Inference in a Nutshell - Cooling House Scenario
2. Causal Structure Learning – Exercises



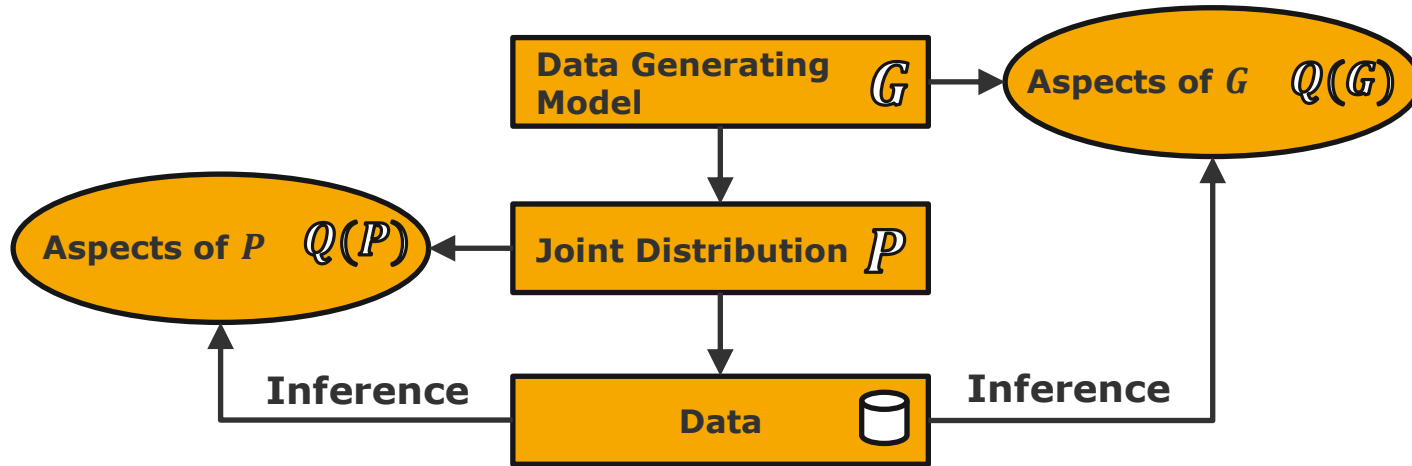
## Recap: Causal Inference in a Nutshell

# Recap: Causal Inference in a Nutshell

## Concept

### Traditional Statistical Inference Paradigm

### Paradigm of Structural Causal Models



E.g., what is the sailors' probability of recovery when **we see** a treatment with lemons?

$$Q(P) = P(\text{recovery}|\text{lemons})$$

E.g., what is the sailors' probability of recovery if **we do** treat them with lemons?

$$Q(G) = P(\text{recovery}|\text{do}(\text{lemons}))$$

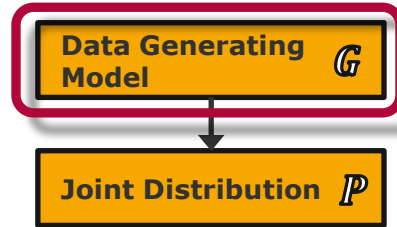
Causal Inference  
Theory and Applications  
in Enterprise Computing

Hagedorn, Huegle,  
Perscheid

Slide 4

# Recap: Causal Inference in a Nutshell

## Causal Graphical Models



### Causal Graphical Model

- *Directed Acyclic Graph (DAG)*  $G = (V, E)$ 
  - *Vertices*  $V_1, \dots, V_n$
  - *Directed edges*  $E = (V_i, V_j)$ , i.e.,  $V_i \rightarrow V_j$
  - *No cycles*
- *Directed Edges* encode direct causes via
  - $V_j = f_j(\text{Pa}(V_j), N_j)$  with independent noise  $N_1, \dots, N_n$

### Causal Sufficiency

- All relevant variables are included in the DAG  $G$

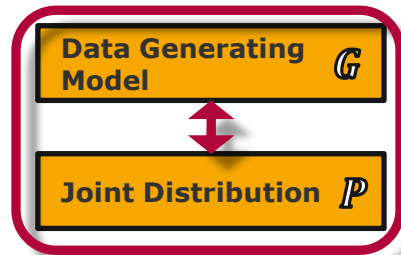
**Causal Inference**  
Theory and Applications  
in Enterprise Computing

Hagedorn, Huegler,  
Perscheid

Slide 5

# Recap: Causal Inference in a Nutshell

## Connecting $G$ and $P$



$$(X \perp\!\!\!\perp Y|Z)_G \Rightarrow (X \perp\!\!\!\perp Y|Z)_P$$

- Key Postulate: *(Local) Markov Condition*
- Essential mathematical concept: *d-Separation*
  - Idea: *Blocking* of paths
  - Implication: *Global Markov Condition*

$$(X \perp\!\!\!\perp Y|Z)_G \Leftarrow (X \perp\!\!\!\perp Y|Z)_P$$

- Key Postulate: *Causal Faithfulness*

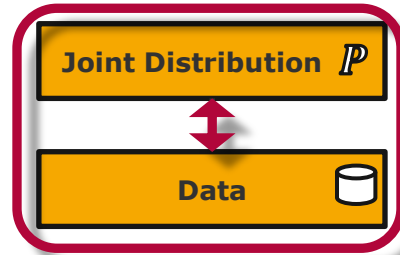
**Causal Inference**  
Theory and Applications  
in Enterprise Computing

Hagedorn, Huegler,  
Perscheid

Slide 6

# Recap: Causal Inference in a Nutshell

Connecting  $P$  and 



## Statistical Inference

- Essential concept: *Point estimator*  $\hat{\theta}$ 
  - *Statistic*  $g(X_1, \dots, X_n)$  of *random samples*  $X_1, \dots, X_n$  to estimate *population parameter*  $\theta$
- Inference: *Statistical Hypothesis Test*
  - *Null Hypothesis*  $H_0$ , claim on a population's property initially assumed to be true
  - *Alternative Hypothesis*  $H_1$ , a claim that contradicts  $H_0$
  - Rejection criteria for  $H_0$ : *c-value*  $T(x) > c$  or equivalently *p-value*  $P_{H_0}(T(X) > T(x)) < \alpha$

$$(X \perp\!\!\!\perp Y|Z)_P \leftarrow \text{cylinder icon}$$

- Key idea: *Conditional Independence Test*
  - Distribution of  $V = \{V_1, \dots, V_N\} \Rightarrow$  dependence measure  $T(V_i, V_j, \mathcal{S}) \Rightarrow$  hypothesis  $H_0: t = 0$

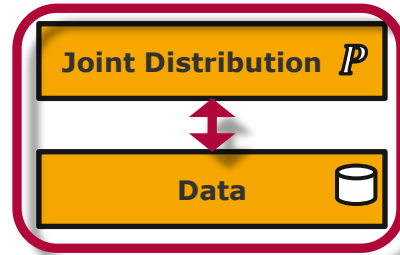
**Causal Inference**  
Theory and Applications  
in Enterprise Computing

Hagedorn, Huegle,  
Perscheid

Slide 7

# Recap: Causal Inference in a Nutshell

Connecting  $P$  and 



## Statistical Inference

- Essential concept: *Point estimator*  $\hat{\theta}$ 
  - *Statistic*  $g(X_1, \dots, X_n)$  of *random samples*  $X_1, \dots, X_n$  to estimate *population parameter*  $\theta$
- Inference: *Statistical Hypothesis Test*
  - *Null Hypothesis*  $H_0$ , claim on a population's property initially assumed to be true
  - *Alternative Hypothesis*  $H_1$ , a claim that contradicts  $H_0$
  - Rejection criteria for  $H_0$ : *c-value*  $T(x) > c$  or equivalently *p-value*  $P_{H_0}(T(X) > T(x)) < \alpha$

**Causal Inference**  
Theory and Applications  
in Enterprise Computing

Hagedorn, Huegle,  
Perscheid

Slide 8

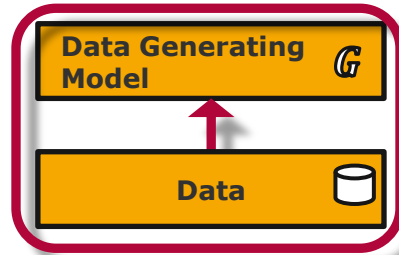
$$(X \perp\!\!\!\perp Y|Z)_P \leftarrow \text{cylinder icon}$$

- Method: *Conditional Independence Test*
  - Distribution of  $V = \{V_1, \dots, V_N\} \Rightarrow$  dependence measure  $T(V_i, V_j, \mathcal{S}) \Rightarrow$  hypothesis  $H_0: t = 0$



# Recap: Causal Inference in a Nutshell

## Causal Structure Learning



### Causal Structure Learning

- Assumptions: *Causal Sufficiency, Markov Condition, Causal Faithfulness*
- Idea: Accept only those DAG's  $G$  for which  $(X \perp\!\!\!\perp Y | Z)_G \Leftrightarrow (X \perp\!\!\!\perp Y | Z)_P$ 
  - Identifies DAG up to *Markov equivalence class* (i.e., same *skeleton*  $C$  and *v-structures*)
  - Markov equivalence class uniquely described by *completed partially directed acyclic graph (CPDAG)*
- Basis:  $V_i$  and  $V_j$  are linked if and only if there is no  $S(V_i, V_j)$  s.t.  $(V_i \perp\!\!\!\perp V_j | S(V_i, V_j))_P$
- Methods:
  - *Constraint-based*: CI testing to derive skeleton together with edge orientation rules
  - *Score-based*: "search-and-score approach"
  - *Hybrid*: Constraint-based skeleton derivation and score-based edge orientation

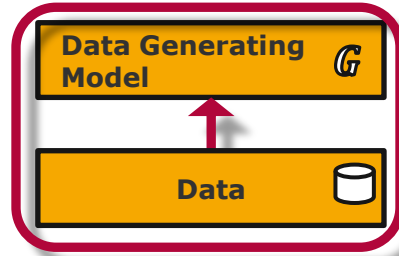
**Causal Inference**  
Theory and Applications  
in Enterprise Computing

Hagedorn, Huegle,  
Perscheid

Slide 9

# Recap: Causal Inference in a Nutshell

## Causal Structure Learning



### PC Algorithm

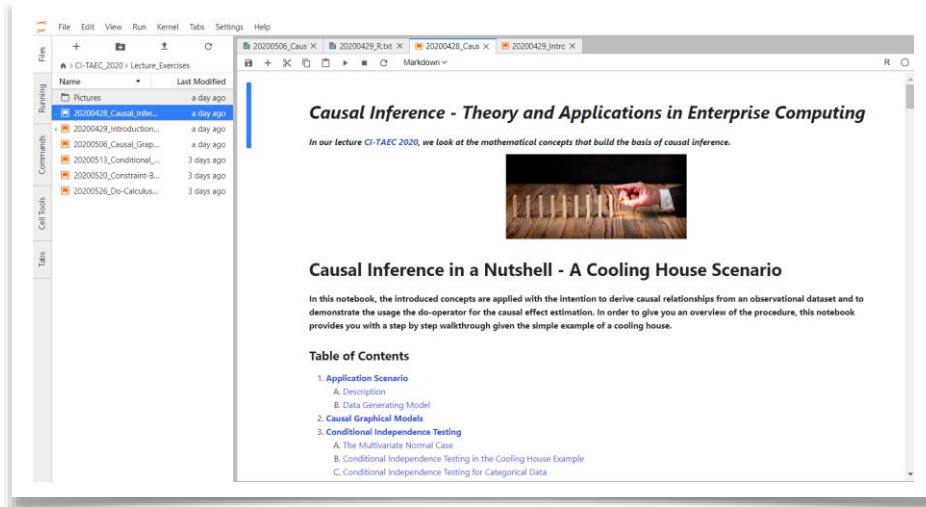
- Concept:
  - *Skeleton discovery*: Iterative CI testing given increasing adjacent  $S(V_i, V_j)$
  - *Edge orientation*: Deterministic orientation rules implied by Markov equivalence class
- Properties:
  - Polynomial complexity given sparse graphs  $G$  (exponential in worst case)
  - Asymptotic consistency (under technical assumptions)  $\Pr(\hat{G} = G) \rightarrow 1 \quad (n \rightarrow \infty)$
  - Extensions allow for weaker faithfulness, latent variables, cycles, etc.

**Causal Inference**  
Theory and Applications  
in Enterprise Computing

Hagedorn, Huegle,  
Perscheid




**Jupyter Lab**



The screenshot shows a Jupyter Lab window with a file explorer on the left and a notebook in the center. The notebook content includes:

### Causal Inference - Theory and Applications in Enterprise Computing

*In our lecture CI-TAEC 2020, we look at the mathematical concepts that build the basis of causal inference.*



### Causal Inference in a Nutshell - A Cooling House Scenario

In this notebook, the introduced concepts are applied with the intention to derive causal relationships from an observational dataset and to demonstrate the usage of the do-operator for the causal effect estimation. In order to give you an overview of the procedure, this notebook provides you with a step by step walkthrough given the simple example of a cooling house.

#### Table of Contents

- 1. Application Scenario
  - A. Description
  - B. Data Generating Model
- 2. Causal Graphical Models
- 3. Conditional Independence Testing
  - A. The Multivariate Normal Case
  - B. Conditional Independence Testing in the Cooling House Example
  - C. Conditional Independence Testing for Categorical Data

## Topics

- CSL in the Cooling House Example

## Causal Inference Theory and Applications in Enterprise Computing

Hagedorn, Huegle,  
Perscheid

# Jupyter Lab

## CI Tests - Exercises

**Table of Contents**

1. Preliminaries
2. Exercises (~40 Minutes)
  - A. Exercises - Fair Coin Tosses and Point Estimator
  - B. Exercises - A Statistical Hypothesis Test for Fairness
  - C. Exercises - Derivation of Independence Test for Two Repeated Coin Tosses
3. Exkurs - Direct Statistical Hypothesis Test

### 1. Preliminaries

Some preliminary libraries required in the provision of functionality to examine the following exercises. Note, that more information about the functionalities of the R-library pcalg can be found in the documentation provided by the CRAN-R project: <https://cran.r-project.org/web/packages/pcalg/pcalg.pdf>.

```
1 | #.install.packages(c("igraph", "Rgraphviz", "pcalg"))
2 | Study1 <- rbinom(n, 1, p_samples[1])
3 | Study2 <- rbinom(n, 1, p_samples[2])
4 | Study3 <- rbinom(n, 1, p_samples[3])
5 | Study4 <- rbinom(n, 1, p_samples[4])
6 | Study5 <- rbinom(n, 1, p_samples[5])
```

### 2. Exercises

Please take **approx 40-50 minutes** to examine the following exercises

#### 2.A. Exercises - Fair Coin Tosses and Point Estimator

We speak of the tossing of a fair coin, if the toss result of this coin has an equal probability of 50% to be either head (1) or number (0), i.e. we assume probability of head is given by  $p_H := P(\text{Coin toss} = 1) = 0.5$  and hence,  $P(\text{Coin toss} = 0) = 1 - P(\text{Coin toss} = 1) = 0.5$ .

---

• Given 100 samples of a fairness study that contain the results of 100 independent coin tosses, what may be a good point estimator  $\hat{p}_H$  to estimate the population parameter  $p_H$ ?

---

Write your assessment and demonstrate its application within our fairness example and the samples in the variable `StudyFair`.  
Assessment:

## Exercises

- Markov Equivalence Class
- Causal Sufficiency
- Causal Faithfulness

## Causal Inference Theory and Applications in Enterprise Computing

Hagedorn, Huegle,  
Perscheid

Slide 13

Thank you  
for your attention!