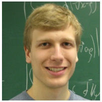
Causal Spaces: A Measure-Theoretic Axiomatisation of Causality

Simon Buchholz[2] Bernhard Schölkopf[2] Krikamol Muandet[3] Fanny Yang[1] Thomas Icard[4]

[1]ETH Zürich

[2]Max Planck Institute for Intelligent Systems, Tübingen

[3]CISPA Helmholtz Center for Information Security, Saarbrücken

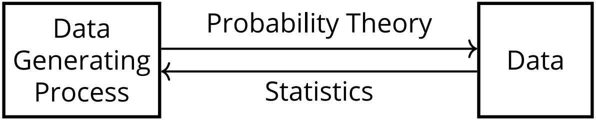
[4]Stanford Univeristy

Spatiotemporal Causality Reading Group, 25 June 2025

Table of Contents

Table of Contents

Probability vs Statistics



Probability Theory

Probability Space

$$(\Omega, \mathcal{H}, \mathbb{P})$$

Set of outcomes → Ω

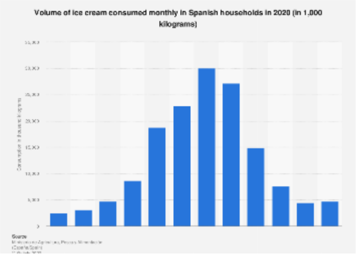
σ-algebra of events → $\mathcal{H}$

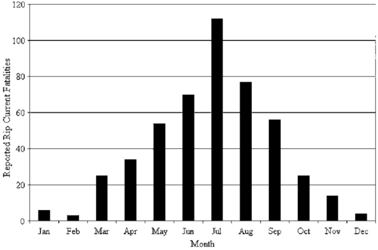
Probability measure → $\mathbb{P}$

Foundations of the Theory of Probability, Andrei N Kolmogorov, 1933.

Is Probability Theory All We Need?

Ice Cream Sales vs Beach Accidents





Is Probability Theory All We Need?

Rain and Plant Growth

Is Probability Theory All We Need?

Is Moderate Drinking Good For Health?

Moderate drinking not better for health than abstaining, analysis suggests

Scientists say flaws in previous research mean health benefits from alcohol were exaggerated

For the regular boozer it is a source of great comfort: the fat pile of studies that say a daily tipple is better for a longer life than avoiding alcohol completely.

Is Probability Theory All We Need?

Is Moderate Drinking Good For Health?

Moderate drinking not better for health than abstaining, analysis suggests

Scientists say flaws in previous research mean health benefits from alcohol were exaggerated

For the regular boozer it is a source of great comfort: the fat pile of studies that say a daily tipple is better for a longer life than avoiding alcohol completely.

But a new analysis challenges the thinking and blames the rosy message on flawed research that compares drinkers with people who are sick and sober.

Scientists in Canada delved into 107 published studies on people's drinking habits and how long they lived. In most cases, they found that drinkers were compared with people who abstained or consumed very little alcohol, without taking into account that some had cut down or quit through ill health.

Why is Causality Important?
Evaluation of Policy / Drug / Business Strategy

Why is Causality Important?

Image Classification

Why is Causality Important?

Large Language Models

Tutorial

Causality for Large Language Models

Zhijing Jin · Sergio Garrido

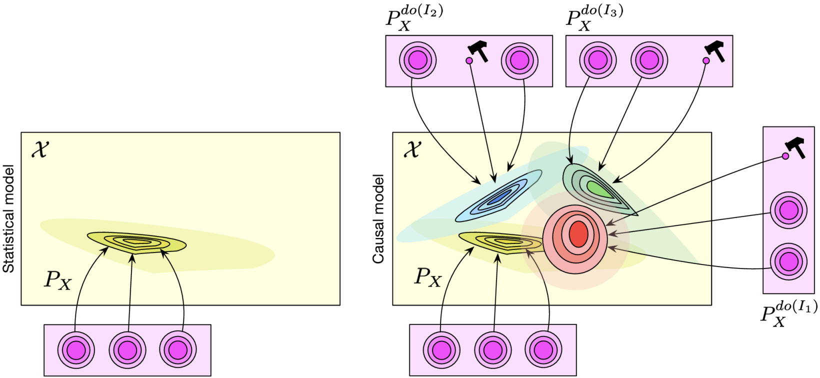
West Ballroom A

[Abstract] [Join Zoom pwd: canstomb]

Tue 10 Dec 9:30 a.m. PST — noon PST (Bookmark)

Please do not share or post zoom links

Manipulation is at the heart of Causality

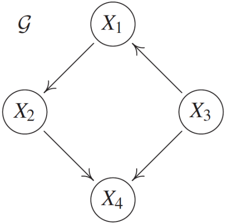


We are interested in what happens to a system, when we intervene on a sub-system.

Schölkopf, Locatello, Bauer, Ke, Kalchbrenner, Goyal and Bengio, *Towards Causal Representation Learning*, 2021.

Existing Frameworks

Structural Causal Models (SCMs)



Potential Outcomes

	$Y^{a=0}$	$Y^{a=1}$
Rheia	0	1
Kronos	1	0
Demeter	0	0
Hades	0	0
Hestia	0	0
Poseidon	1	0
Hera	0	0
Zeus	0	1
Artemis	1	1
Apollo	1	0

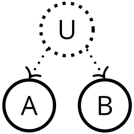
Causality, Pearl, Cambridge University Press, 2009

Elements of Causal Inference, Peters, Janzing and Schölkopf, MIT Press, 2017

Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction, Rubin and Imbens, Cambridge University Press, 2015

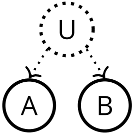
Limitations of Existing Frameworks

- Latent confounding

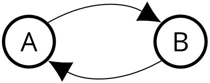


Limitations of Existing Frameworks

- Latent confounding

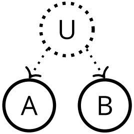


- Cyclic causal relationships

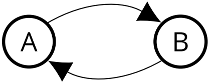


Limitations of Existing Frameworks

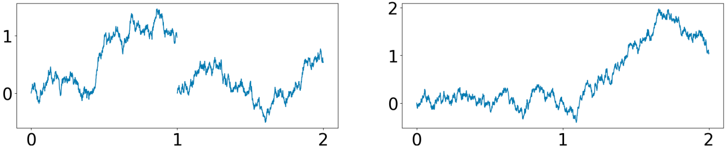
- Latent confounding



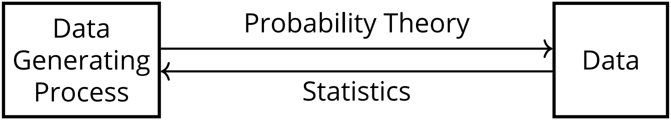
- Cyclic causal relationships



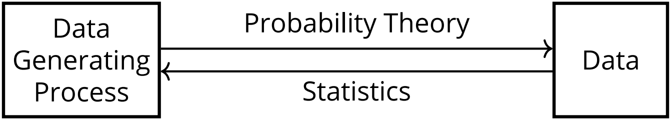
- Continuous-time stochastic processes

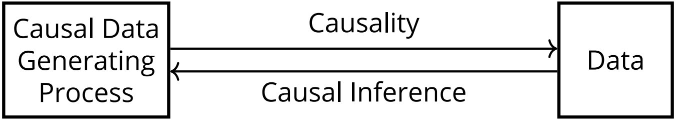


Probability vs Statistics

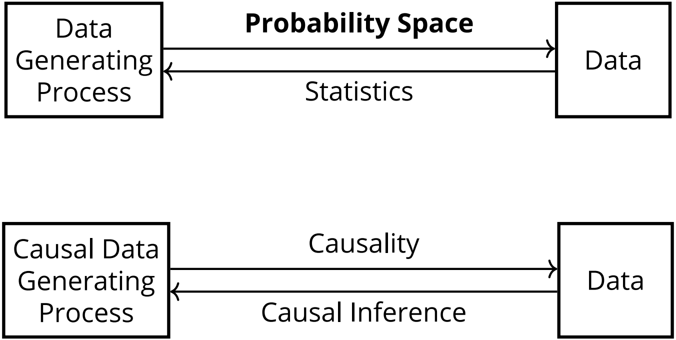


Probability vs Statistics

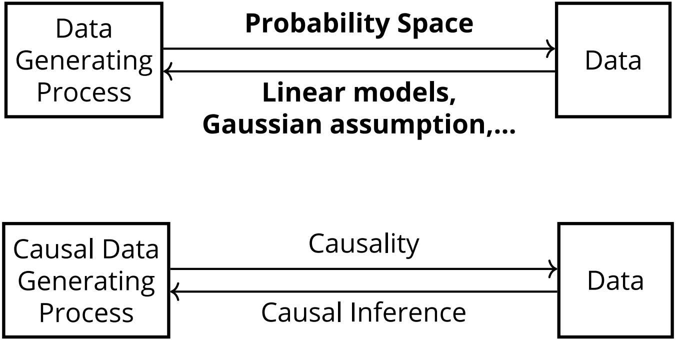




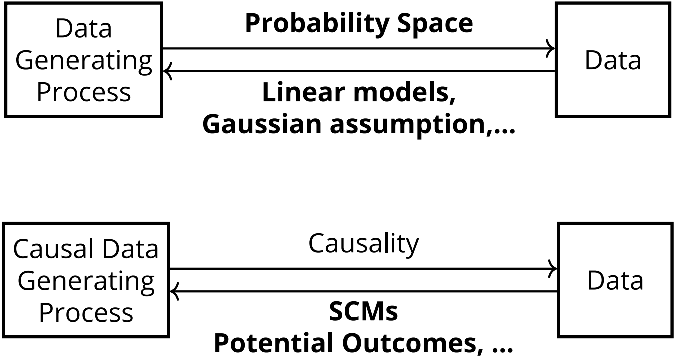
Probability vs Statistics



Probability vs Statistics



Probability vs Statistics



Probability vs Statistics

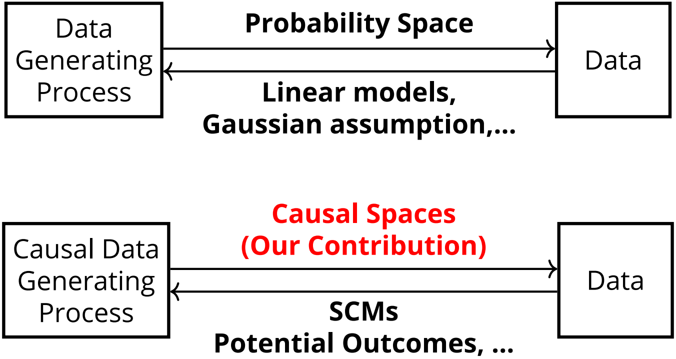


Table of Contents

Notations

- For a set T, we denote its power set by $\mathcal{P}(T)$.

Notations

- For a set T, we denote its power set by $\mathcal{P}(T)$.
- Product measurable space with index set T:

$$(\Omega, \mathcal{H}) = (\times_{t \in T} E_t, \otimes_{t \in T} \mathcal{E}_t).$$

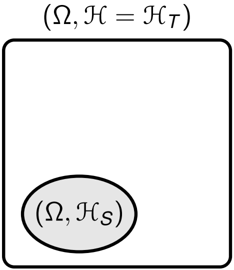
$\mathcal{H}_S$: sub-σ-algebra of $\mathcal{H}$ corresponding to $S \in \mathcal{P}(T)$.

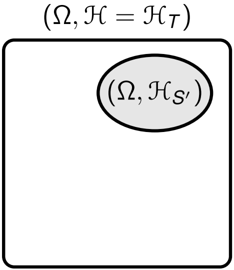
Notations

- For a set T, we denote its power set by $\mathcal{P}(T)$.
- Product measurable space with index set T:

$$(\Omega, \mathcal{H}) = (\times_{t \in T} E_t, \otimes_{t \in T} \mathcal{E}_t).$$

$\mathcal{H}_S$: sub-σ-algebra of $\mathcal{H}$ corresponding to $S \in \mathcal{P}(T)$.
Intuition: $\mathcal{H} = \mathcal{H}_T$ is the entire space. $\mathcal{H}_S$ is a subspace.





Notations

- For a set T, we denote its power set by $\mathcal{P}(T)$.
- Product measurable space with index set T:

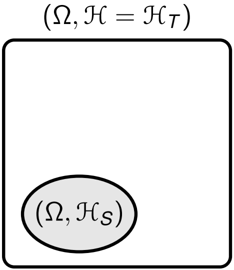
$$(\Omega, \mathcal{H}) = (\times_{t \in T} E_t, \otimes_{t \in T} \mathcal{E}_t).$$

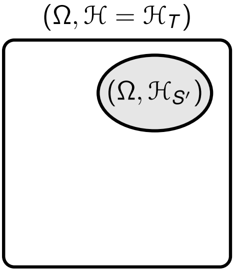
  $\mathcal{H}_S$: sub-σ-algebra of $\mathcal{H}$ corresponding to $S \in \mathcal{P}(T)$.
  Intuition: $\mathcal{H} = \mathcal{H}_T$ is the entire space. $\mathcal{H}_S$ is a subspace.

- "Transition probability kernel"
  K_S from $(\Omega, \mathcal{H}_S)$ into $(\Omega, \mathcal{H})$:

$$K_S(x, \cdot) \to [0, 1].$$

  For every $x \in (\Omega, \mathcal{H}_S)$, $K_S(x, \cdot)$ is a measure on $(\Omega, \mathcal{H})$.





Notations

- For a set T, we denote its power set by $\mathcal{P}(T)$.
- Product measurable space with index set T:

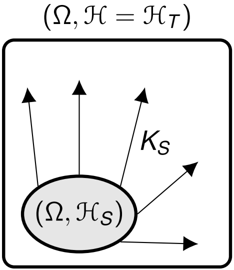
$$(\Omega, \mathcal{H}) = (\times_{t \in T} E_t, \otimes_{t \in T} \mathcal{E}_t).$$

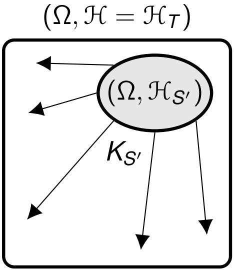
  $\mathcal{H}_S$: sub-σ-algebra of $\mathcal{H}$ corresponding to $S \in \mathcal{P}(T)$.
  Intuition: $\mathcal{H} = \mathcal{H}_T$ is the entire space. $\mathcal{H}_S$ is a subspace.

- "Transition probability kernel"
  K_S from $(\Omega, \mathcal{H}_S)$ into $(\Omega, \mathcal{H})$:

$$K_S(x, \cdot) \to [0, 1].$$

  For every $x \in (\Omega, \mathcal{H}_S)$, $K_S(x, \cdot)$ is a measure on $(\Omega, \mathcal{H})$.
  Intuition: conditional distribution.





Causal Spaces

A **causal space** is a quadruple $(\Omega, \mathcal{H}, \mathbb{P}, \mathbb{K})$, where

$$(\Omega, \mathcal{H}, \mathbb{P}) = (\times_{t \in T} E_t, \otimes_{t \in T} \mathcal{E}_t, \mathbb{P});$$

and

$$\mathbb{K} = \{K_S : S \subseteq T\}, \qquad K_S : (\Omega, \mathcal{H}_S) \to (\Omega, \mathcal{H})$$

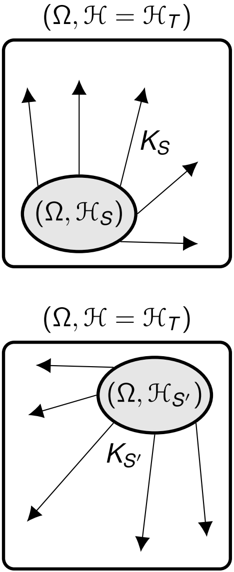
where K_S are **causal kernels**, such that

(i) for all $A \in \mathcal{H}$ and $x \in \Omega$,

$$K_\emptyset(x, A) = \mathbb{P}(A);$$

(ii) for all $A \in \mathcal{H}_S$ and $x \in \Omega$,

$$K_S(x, A) = 1_A(x).$$



Causal Spaces

A **causal space** is a quadruple $(\Omega, \mathcal{H}, \mathbb{P}, \mathbb{K})$, where

$$(\Omega, \mathcal{H}, \mathbb{P}) = (\times_{t \in T} E_t, \otimes_{t \in T} \mathcal{E}_t, \mathbb{P});$$

and

$$\mathbb{K} = \{K_S : S \subseteq T\}, \qquad K_S : (\Omega, \mathcal{H}_S) \to (\Omega, \mathcal{H})$$

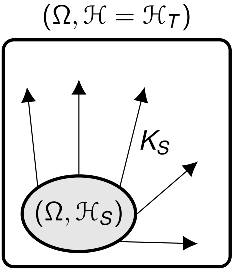
where K_S are **causal kernels**, such that

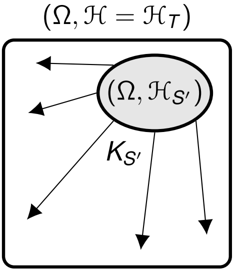
(i) for all $A \in \mathcal{H}$ and $x \in \Omega$,

$$K_\emptyset(x, A) = \mathbb{P}(A);$$

(ii) for all $A \in \mathcal{H}_S$ and $x \in \Omega$,

$$K_S(x, A) = 1_A(x).$$





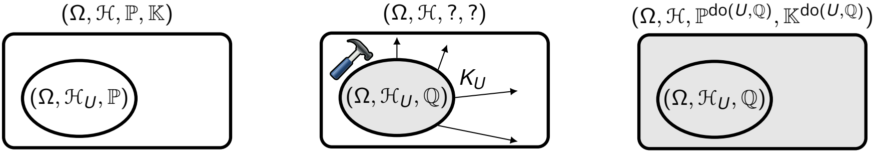
$\mathbb{P}$ is the "observational distribution".

Interventions

An intervention is the process of

(a) choosing a sub-σ-algebra $\mathcal{H}_U$, and

(b) placing any measure $\mathbb{Q}$ on $(\Omega, \mathcal{H}_U)$.



Interventions

An intervention is the process of

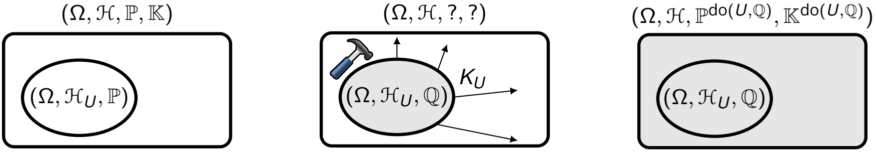
(a) choosing a sub-σ-algebra $\mathcal{H}_U$, and
(b) placing any measure $\mathbb{Q}$ on $(\Omega, \mathcal{H}_U)$.

New causal space: $(\Omega, \mathcal{H}, \mathbb{P}^{\text{do}(U,\mathbb{Q})}, \mathbb{K}^{\text{do}(U,\mathbb{Q})})$, where

$$\mathbb{P}^{\text{do}(U,\mathbb{Q})}(A) = \int \mathbb{Q}(d\omega) K_U(\omega, A)$$

and $\mathbb{K}^{\text{do}(U,\mathbb{Q})} = \{K_S^{\text{do}(U,\mathbb{Q})} : S \in \mathcal{P}(T)\}$ with

$$K_S^{\text{do}(U,\mathbb{Q})}(\omega, A) = \int \mathbb{Q}(d\omega'_{U \setminus S}) K_{S \cup U}((\omega_S, \omega'_{U \setminus S}), A).$$



Axioms of Causal Kernels

Recall that causal kernels K_S from $(\Omega, \mathcal{H}_S)$ into $(\Omega, \mathcal{H})$ satisfy

(i) for all $A \in \mathcal{H}$ and $x \in \Omega$,

$$K_\emptyset(x, A) = \mathbb{P}(A);$$

(ii) for all $A \in \mathcal{H}_S$ and $x \in \Omega$,

$$K_S(x, A) = 1_A(x).$$

Intuition on the axioms:

(i) $\mathbb{P}^{\mathrm{do}(\emptyset,\mathbb{Q})}(A) = \mathbb{P}(A)$.

(ii) For $A \in \mathcal{H}_U$, $\mathbb{P}^{\mathrm{do}(U,\mathbb{Q})}(A) = \int \mathbb{Q}(dx) 1_A(x) = \mathbb{Q}(A)$.

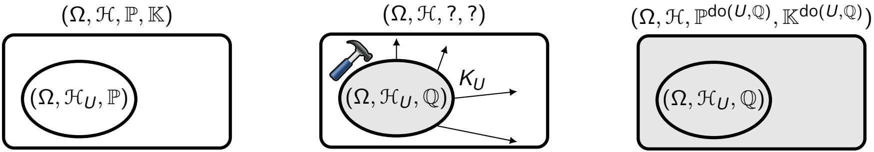
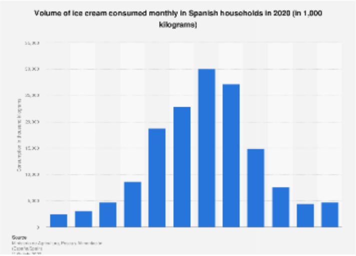
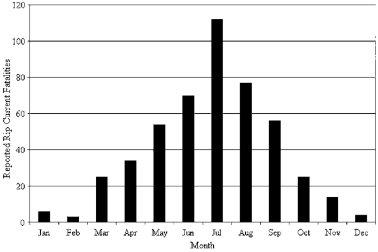


Table of Contents

Ice Cream Sales and Beach Accidents

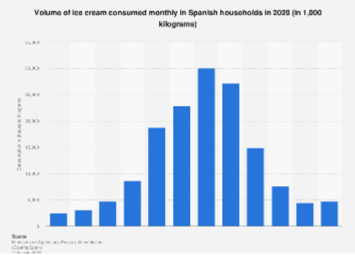


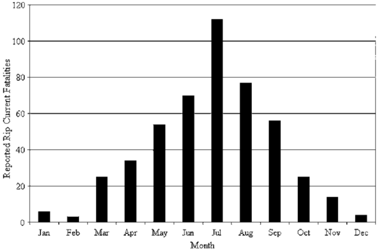


- Causal space: $(E_{\text{ice}} \times E_{\text{acc}}, \mathcal{E}_{\text{ice}} \otimes \mathcal{E}_{\text{acc}}, \mathbb{P}, \mathbb{K})$[1].

[1] $E_{\text{ice}} = E_{\text{acc}} = \mathbb{R}$ and $\mathcal{E}_{\text{ice}} = \mathcal{E}_{\text{acc}}$ is the Lebesgue σ-algebra.

Ice Cream Sales and Beach Accidents

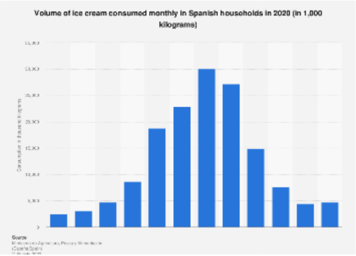


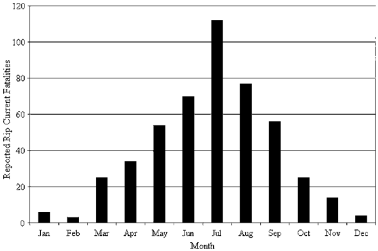


- Causal space: $(E_{\text{ice}} \times E_{\text{acc}}, \mathcal{E}_{\text{ice}} \otimes \mathcal{E}_{\text{acc}}, \mathbb{P}, \mathbb{K})$[1].
- $\mathbb{P}$ has strong correlation.

[1] $E_{\text{ice}} = E_{\text{acc}} = \mathbb{R}$ and $\mathcal{E}_{\text{ice}} = \mathcal{E}_{\text{acc}}$ is the Lebesgue σ-algebra.

Ice Cream Sales and Beach Accidents





- Causal space: $(E_{\text{ice}} \times E_{\text{acc}}, \mathcal{E}_{\text{ice}} \otimes \mathcal{E}_{\text{acc}}, \mathbb{P}, \mathbb{K})$[1].
- $\mathbb{P}$ has strong correlation.
- For causal kernels, let
  - $K_{\text{ice}}(x, A) = \mathbb{P}(A)$ for all $A \in \mathcal{E}_{\text{acc}}$; and
  - $K_{\text{acc}}(y, B) = \mathbb{P}(B)$ for all $B \in \mathcal{E}_{\text{ice}}$.

[1] $E_{\text{ice}} = E_{\text{acc}} = \mathbb{R}$ and $\mathcal{E}_{\text{ice}} = \mathcal{E}_{\text{acc}}$ is the Lebesgue σ-algebra.

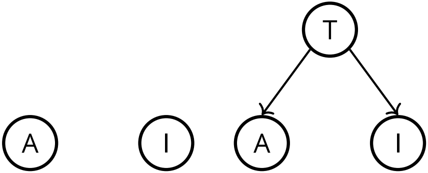
Ice Cream Sales and Beach Accidents

A Number of beach accidents.

I Ice cream sales.

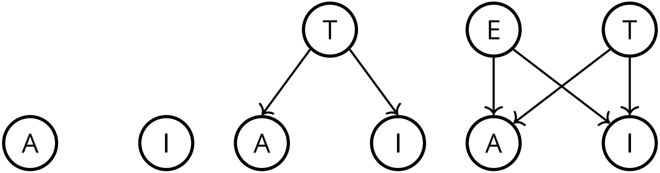
Ice Cream Sales and Beach Accidents

- A Number of beach accidents.
- I Ice cream sales.
- T Temperature.



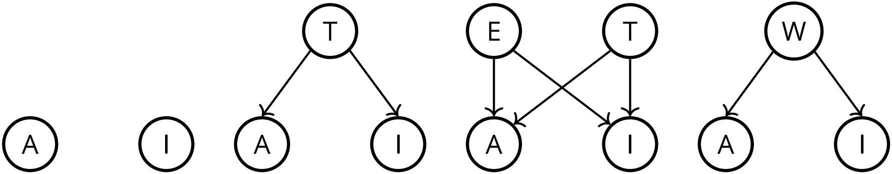
Ice Cream Sales and Beach Accidents

- A Number of beach accidents.
- I Ice cream sales.
- T Temperature.
- E Economy.



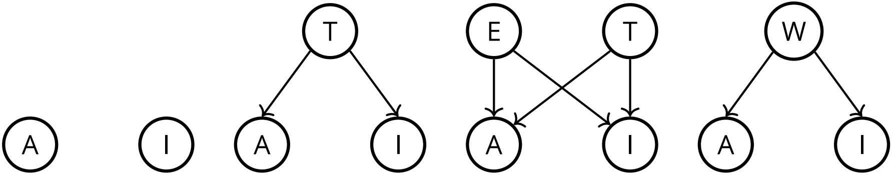
Ice Cream Sales and Beach Accidents

- A Number of beach accidents.
- I Ice cream sales.
- T Temperature.
- E Economy.
- W World.



Ice Cream Sales and Beach Accidents

- A Number of beach accidents.
- I Ice cream sales.
- T Temperature.
- E Economy.
- W World.



Recall:

- Causal space: $(E_{\text{ice}} \times E_{\text{acc}}, \mathcal{E}_{\text{ice}} \otimes \mathcal{E}_{\text{acc}}, \mathbb{P}, \mathbb{K})$.
- $\mathbb{P}$ has strong correlation.
- For causal kernels, let
  - $K_{\text{ice}}(x, A) = \mathbb{P}(A)$ for all $A \in \mathcal{E}_{\text{acc}}$; and
  - $K_{\text{acc}}(y, B) = \mathbb{P}(B)$ for all $B \in \mathcal{E}_{\text{ice}}$.

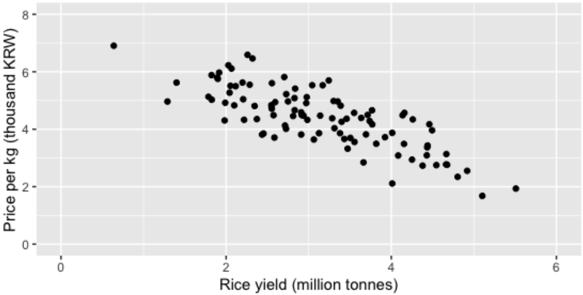
Crop Yield and Price

- Causal space: $(E_{\text{rice}} \times E_{\text{price}}, \mathcal{E}_{\text{rice}} \otimes \mathcal{E}_{\text{price}}, \mathbb{P}, \mathbb{K})$[2].

[2] $E_{\text{rice}} = E_{\text{price}} = \mathbb{R}$ and $\mathcal{E}_{\text{rice}} = \mathcal{E}_{\text{price}}$ is the Lebesgue σ-algebra.

Crop Yield and Price

- Causal space: $(E_{\text{rice}} \times E_{\text{price}}, \mathcal{E}_{\text{rice}} \otimes \mathcal{E}_{\text{price}}, \mathbb{P}, \mathbb{K})^2$.
- No Intervention:



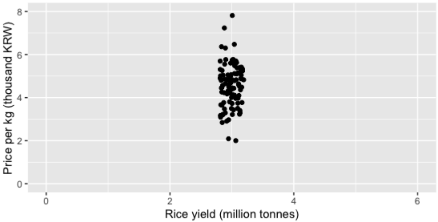
[2] $E_{\text{rice}} = E_{\text{price}} = \mathbb{R}$ and $\mathcal{E}_{\text{rice}} = \mathcal{E}_{\text{price}}$ is the Lebesgue σ-algebra.

Crop Yield and Price

- Intervention to fix rice $= 3$:

$$K_{\text{rice}}(3, A) = \int_A \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-4.5)^2} dx$$

for $A \in \mathcal{E}_{\text{price}}$.

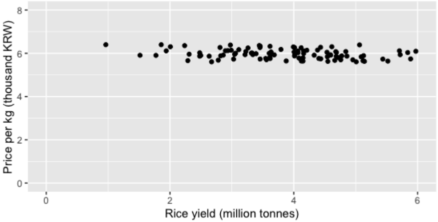


Crop Yield and Price

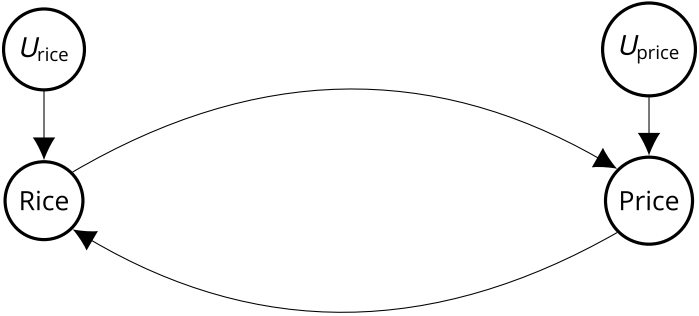
- Intervention to fix price $= 6$:

$$K_{\text{price}}(6, B) = \int_B \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-4)^2} dx$$

for $B \in \mathcal{E}_{\text{rice}}$.

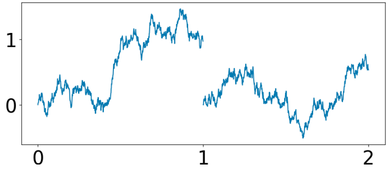


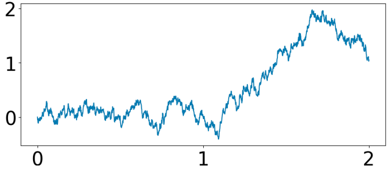
Crop Yield and Price



$$\text{Rice} = f_{\text{rice}}(\text{Price}, U_{\text{rice}}), \qquad \text{Price} = f_{\text{price}}(\text{Rice}, U_{\text{price}}).$$

1-dimensional Brownian Motion

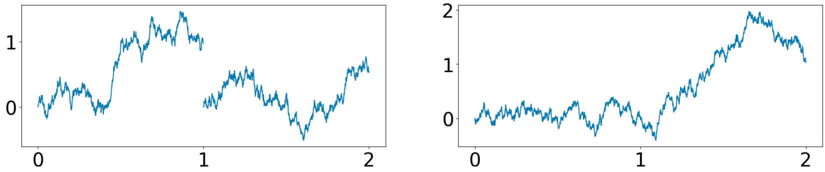




- Causal space: $(\times_{t\in\mathbb{R}_+} E_t, \otimes_{t\in\mathbb{R}_+} \mathcal{E}_t, \mathbb{P}, \mathbb{K})$[3].

[3] For each $t \in \mathbb{R}_+$, $E_t = \mathbb{R}$ and $\mathcal{E}_t$ is the Lebesgue σ-algebra.

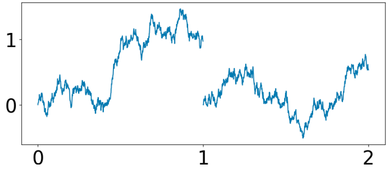
1-dimensional Brownian Motion

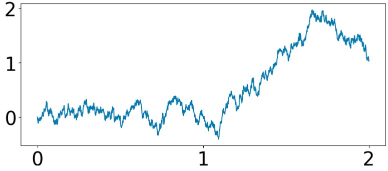


- Causal space: $(\times_{t\in\mathbb{R}_+} E_t, \otimes_{t\in\mathbb{R}_+} \mathcal{E}_t, \mathbb{P}, \mathbb{K})^3$.
- $\mathbb{P}$ is the Wiener measure.

[3] For each $t \in \mathbb{R}_+$, $E_t = \mathbb{R}$ and $\mathcal{E}_t$ is the Lebesgue σ-algebra.

1-dimensional Brownian Motion



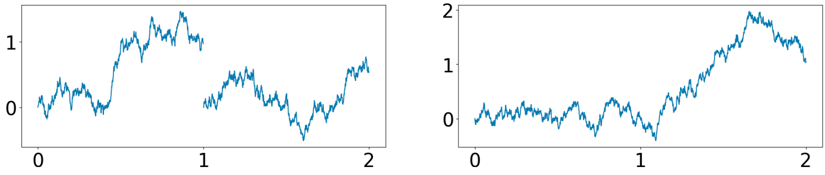


- Causal space: $(\times_{t\in\mathbb{R}_+} E_t, \otimes_{t\in\mathbb{R}_+} \mathcal{E}_t, \mathbb{P}, \mathbb{K})$[3].
- $\mathbb{P}$ is the Wiener measure.
- For any $s < t$, the causal kernels are

$$K_s(x, y) = \frac{1}{\sqrt{2\pi(t-s)}} e^{-\frac{1}{2(t-s)}(y-x)^2}, \qquad K_t(x, y) = \frac{1}{\sqrt{2\pi s}} e^{-\frac{1}{2s} y^2}.$$

[3] For each $t \in \mathbb{R}_+$, $E_t = \mathbb{R}$ and $\mathcal{E}_t$ is the Lebesgue σ-algebra.

1-dimensional Brownian Motion



- Causal space: $(\times_{t\in\mathbb{R}_+} E_t, \otimes_{t\in\mathbb{R}_+} \mathcal{E}_t, \mathbb{P}, \mathbb{K})^3$.
- $\mathbb{P}$ is the Wiener measure.
- For any $s < t$, the causal kernels are

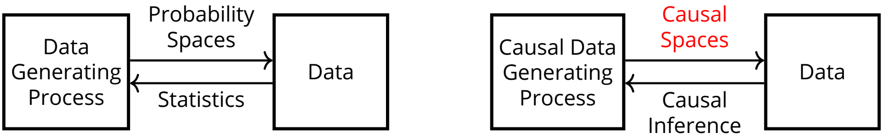
$$K_s(x,y) = \frac{1}{\sqrt{2\pi(t-s)}} e^{-\frac{1}{2(t-s)}(y-x)^2}, \qquad K_t(x,y) = \frac{1}{\sqrt{2\pi s}} e^{-\frac{1}{2s}y^2}.$$

Past values affect the future, but future values do not affect the past.

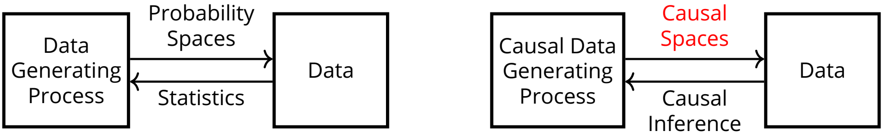
[3] For each $t \in \mathbb{R}_+$, $E_t = \mathbb{R}$ and $\mathcal{E}_t$ is the Lebesgue σ-algebra.

Table of Contents

Summary


Data Generating Process
Probability Spaces
Statistics
Data
Causal Data Generating Process
Causal Spaces
Causal Inference
Data

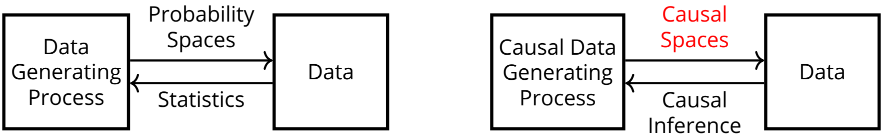
Summary



Causal spaces

- are a **measure-theoretic axiomatisation** of causality;

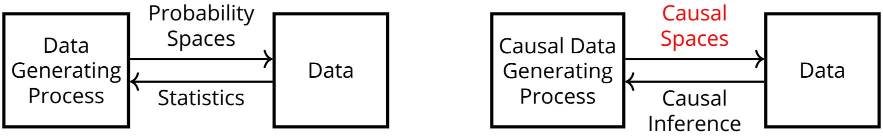
Summary



Causal spaces

- are a **measure-theoretic axiomatisation** of causality;
- strictly generalise existing frameworks;

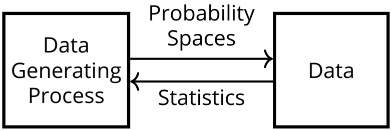
Summary

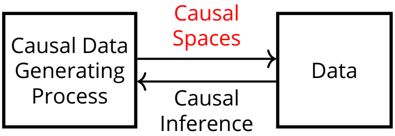


Causal spaces

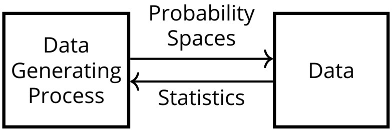
- are a **measure-theoretic axiomatisation** of causality;
- strictly generalise existing frameworks;
- overcome long-standing limitations of existing frameworks, for example,
  - latent confounders,
  - cyclic causal relationships,
  - continuous-time stochastic processes.

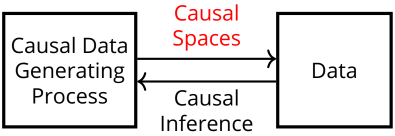
Summary


Data Generating Process
Probability Spaces
Statistics
Data


Causal Data Generating Process
Causal Spaces
Causal Inference
Data

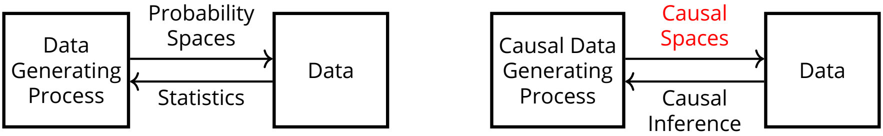
Summary





- We hope that causal spaces can play the same role as probability spaces for causality.

Summary



- We hope that causal spaces can play the same role as probability spaces for causality.
- Our view is that SCMs and potential outcomes are valuable tools to specify a causal space.

Not covered in this talk

- Causal effect [1].
- Sources & identifiability [1].
- Causal stochastic processes [1].

[1] A Measure-Theoretic Axiomatisation of Causality, P., Buchholz, Schölkopf and Muandet, NeurIPS 2023

[2] Products, Abstractions and Inclusions of Causal Spaces, Buchholz*, P.* and Schölkopf, UAI 2024

Not covered in this talk

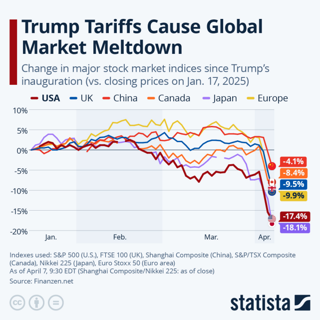
- Causal effect [1].
- Sources & identifiability [1].
- Causal stochastic processes [1].
- Multiple causal spaces [2].
- Causal independence [2].

[1] A Measure-Theoretic Axiomatisation of Causality, P., Buchholz, Schölkopf and Muandet, NeurIPS 2023
[2] Products, Abstractions and Inclusions of Causal Spaces, Buchholz*, P.* and Schölkopf, UAI 2024

Not covered in this talk

Counterfactuals

- Counterfactuals [3].
- Fundamental Theorem of Causality [3].
- Independence and Synchronisation of counterfactual worlds [3].



[3] Counterfactual Causal Spaces and the Fundamental Theorem of Causality, P., Yang and Icard, to be submitted soon.

Thank you.