A Measure-Theoretic Axiomatisation of Causality

Junhyung Park[1], Simon Buchholz[1], Bernhard Schölkopf[1], Krikamol Muandet[2]
[1]Max Planck Institute for Intelligent Systems, Tübingen, [2]CISPA Helmholtz Center for Information Security, Saarbrücken

Overview

- Causality is a central concept in many research domains, and there are many mathematical frameworks that encode causal information, but there is no universally accepted axiomatisation of it.
- We view the concept of causality both as an extension of probability theory, and as a study of *what happens to a system when we manipulate on a subsystem.*
- Based on these two main ideas, we propose *causal spaces* as an axiomatic framework of causality.
- Causal spaces not only strictly generalise existing frameworks, but it also sheds light on long-standing limitations of existing frameworks (e.g. structural causal models (SCMs) or the potential outcomes framework) including, for example cyclic causal relationships, latent variables and continuous-time stochastic processes.

Causality as an extension of Probability Theory

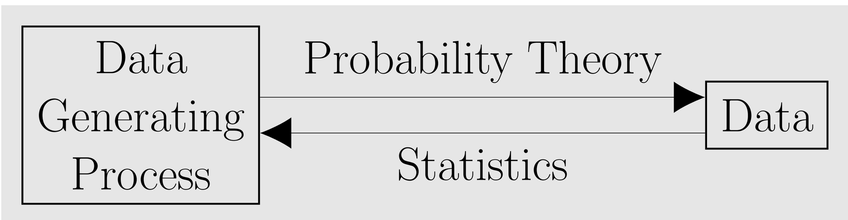


Fig. 1: Statistics is an inverse problem of probability theory.

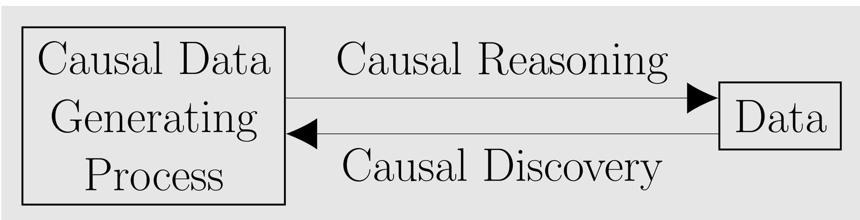


Fig. 2: Causal discovery is an inverse problem of reasoning.

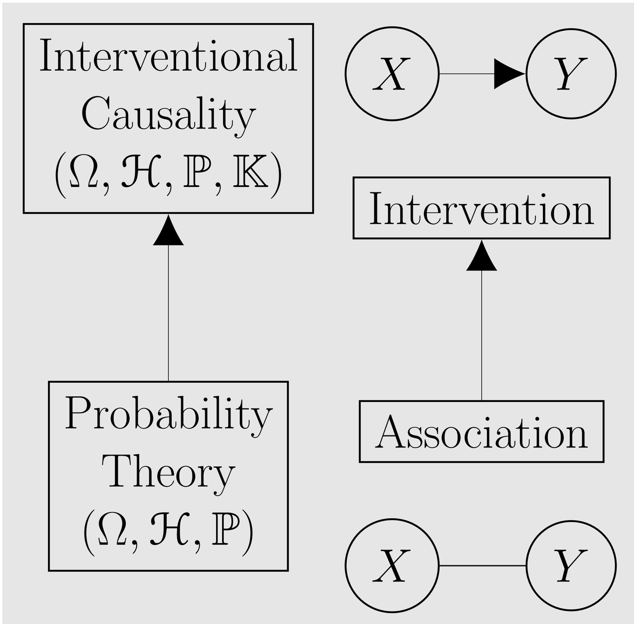


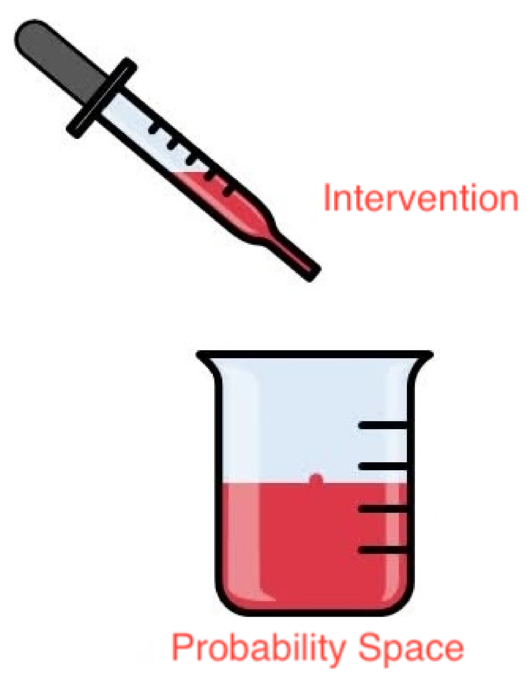
Fig. 3: Pearl's ladder of causation (bottom two rungs)

Causal Spaces

Let T be any index set, $\mathcal{P}(T)$ its power set, and for $S\subseteq T$, let $\mathcal{H}_S$ be the sub-σ-algebra of $\mathcal{H}$ corresponding to S.
A *causal space* is the quadruple $(\Omega,\mathcal{H},\mathbb{P},\mathbb{K})$, where:

- $(\Omega,\mathcal{H},\mathbb{P})=(\times_{t\in T}E_t,\otimes_{t\in T}\mathcal{E}_t,\mathbb{P})$ is a probability space;
- $\mathbb{K}=\{K_S:S\in\mathcal{P}(T)\}$ is a collection of transition probability kernels K_S from $(\Omega,\mathcal{H}_S)$ into $(\Omega,\mathcal{H})$, called the *causal kernel on* $\mathcal{H}_S$, such that
  (i) for all $A\in\mathcal{H}$ and $\omega\in\Omega$,
  $$K_\emptyset(\omega,A)=\mathbb{P}(A);$$
  (ii) for all $A\in\mathcal{H}_S$ and $\omega\in\Omega$,
  $$K_S(\omega,A)=1_A(\omega).$$

$\mathbb{P}$ is the observational measure.



Interventions

Intervention is the process of choosing a sub-σ-algebra $\mathcal{H}_U$ and placing any measure $\mathbb{Q}$ on $(\Omega,\mathcal{H}_U)$. Then we have a new causal space $(\Omega,\mathcal{H},\mathbb{P}^{\text{do}(U,\mathbb{Q})},\mathbb{K}^{\text{do}(U,\mathbb{Q})})$, where

$$\mathbb{P}^{\text{do}(U,\mathbb{Q})}(A)=\int\mathbb{Q}(d\omega)K_U(\omega,A)\tag{1}$$

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

$$K_S^{\text{do}(U,\mathbb{Q})}(\omega,A)=\int\mathbb{Q}(d\omega'_{U\setminus S})K_{S\cup U}((\omega_S,\omega'_{U\setminus S}),A).\tag{2}$$

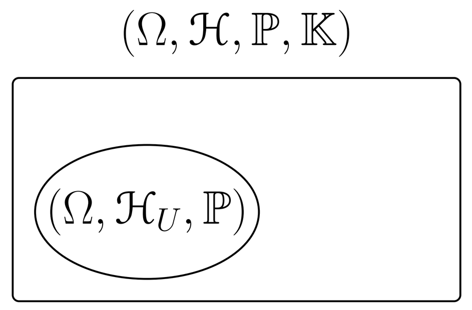


Fig. 5: Observational state.

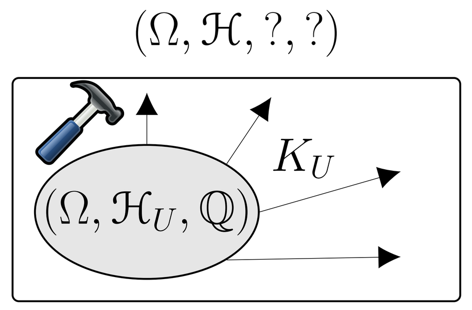


Fig. 6: Intervention.

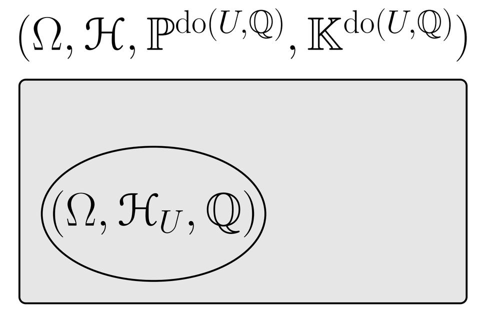


Fig. 7: New state after intervention.

Intuition on the axioms

(i) Intervening on nothing leaves the measure intact, $\mathbb{P}^{\text{do}(\emptyset,\mathbb{Q})}(A)=\mathbb{P}(A)$.

(ii) Intervening on a sub-σ-algebra returns a measure which, when restricted to that sub-σ-algebra, is precisely the measure we give it, $\mathbb{P}^{\text{do}(U,\mathbb{Q})}(A)=\mathbb{Q}(A)$ for $A\in\mathcal{H}_U$.

How is the causal information encoded?

SCMs: $X=f_X(\ldots)$

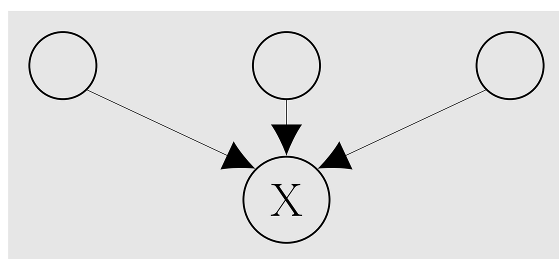


Fig. 8: The encoded causal information is *how the system affects a sub-system.*

Causal Spaces: $K_X(x,\ldots)=\ldots$

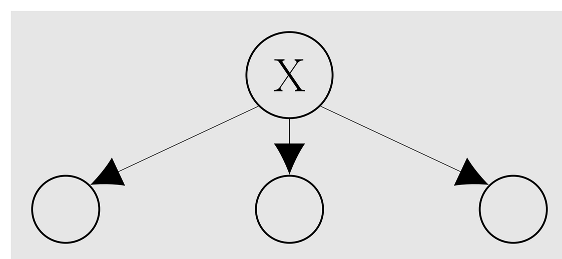


Fig. 9: The encoded causal information is *how a sub-system affects the system.*

What are the primitive objects?

SCMs

- Set of variables $X_1,\ldots,X_n$.
- Noise distribution.
- Structural equations $X_i=f_i(\ldots)$.
- The system is *deterministic.*
- Observational and interventional distributions are *derived* from these objects.

Causal Spaces

- Probability space $(\Omega,\mathcal{H},\mathbb{P})$.
- Causal kernels K_S.
- The system is *stochastic.*
- Observational (via $\mathbb{P}$) and interventional distributions (via the causal kernels) *are* the primitive objects.

Example: Latent Variables

Causal Space: $(E_{\text{ice}}\times E_{\text{acc}},\mathcal{E}_{\text{ice}}\otimes\mathcal{E}_{\text{acc}},\mathbb{P},\{K_\emptyset,K_{\text{ice}},K_{\text{acc}},K_{\text{ice,acc}}\})$. $\mathbb{P}$ has correlation.

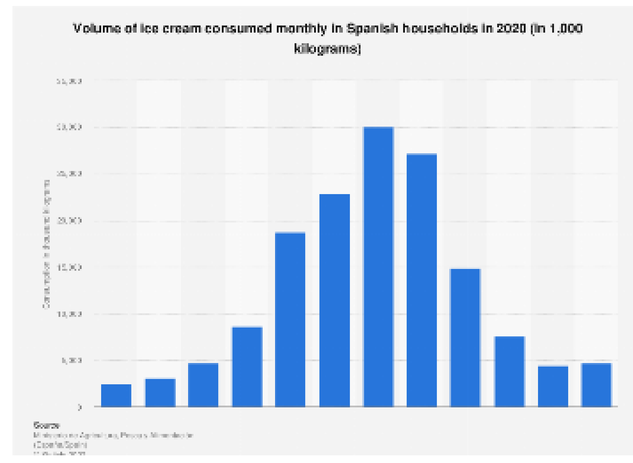


Fig. 10: Ice cream sales by month

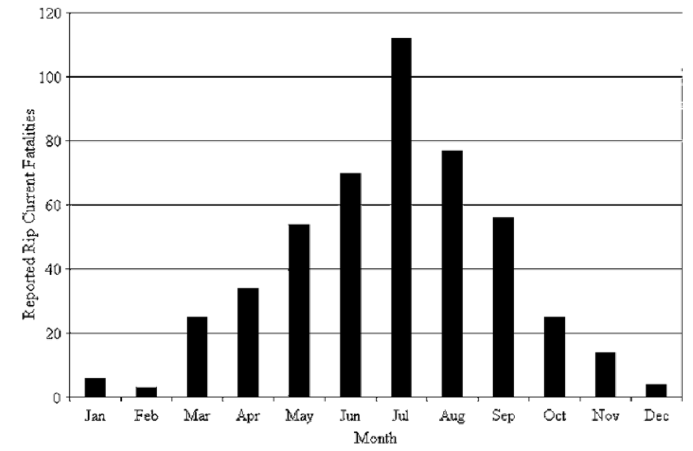


Fig. 11: Fatal rip current accidents by month

$K_{\text{ice}}(x,A)=\mathbb{P}(A)$ for all $A\in\mathcal{E}_{\text{acc}}$ and $K_{\text{acc}}(y,B)=\mathbb{P}(B)$ for all $B\in\mathcal{E}_{\text{ice}}$.

Example: Cyclic Causal Relationship

Causal Space: $(E_{\text{rice}}\times E_{\text{price}},\mathcal{E}_{\text{rice}}\otimes\mathcal{E}_{\text{price}},\mathbb{P},\{K_\emptyset,K_{\text{rice}},K_{\text{price}},K_{\text{rice,price}}\})$.

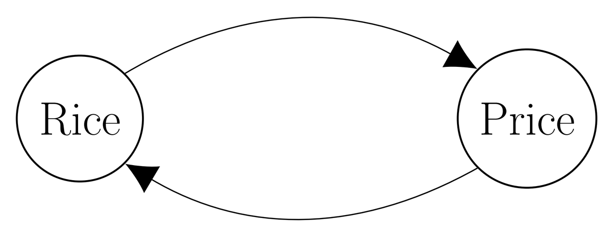


Fig. 12: Cyclic Causal Relationship

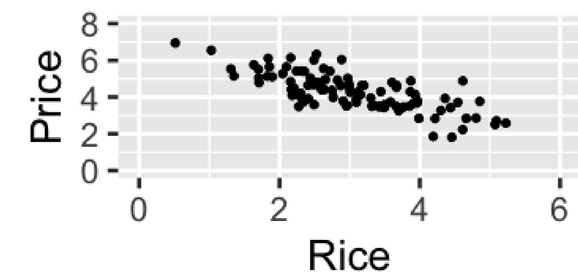


Fig. 14: Observational Measure, $\mathbb{P}$

Fig. 13: $K_{\text{rice}}(3,p)=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(p-4.5)^2}$.

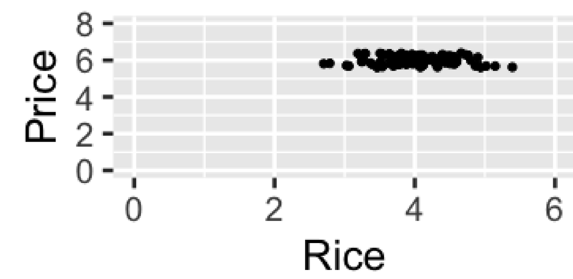


Fig. 15: $K_{\text{price}}(6,r)=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(r-4)^2}$.

Example: Continuous-Time Stochastic Process

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

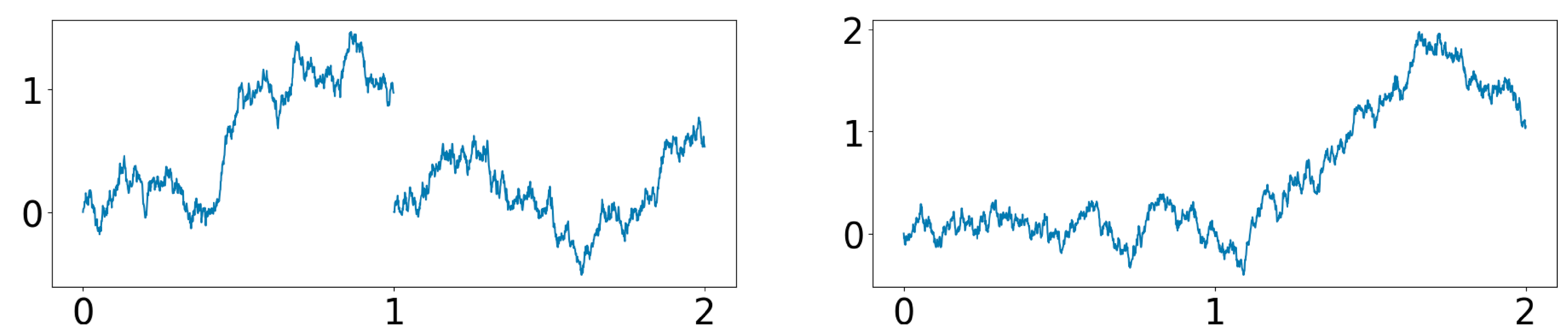


Fig. 16: The left plot shows intervening at time 1, whereas the right plot shows conditioning at time 1.

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

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