A Measure-Theoretic Axiomatisation of Causality (NeurIPS, 2023)

SeongSik Choi

January 8, 2024

Seoul National University

Introduction

이 논문은 causality 이론들을 측도론(measure-theory)을 이용하여 공리화(axiomatisation)하는 framework를 제안한 논문입니다.

기존의 Probability space에 causal mechanism이라 불리는 요소를 추가하여 causal space라는 개념을 정의하게 됩니다.

Causal space는 기존의 주요한 두 인과 추론 framework인 Stochastic Causal Models(SCMs)와 Potential Outcomes(PO) framework를 잘 encode할 수 있으며,

또한 cycle이 있는 경우와 continuous-time stochastic processes 등, 기존 framework를 이용하여 수학적으로 표현되기 어려운 상황들도 잘 표현할 수 있습니다.

Transition probability kernel

이 논문은 causality 이론들을 측도론(measure-theory)을 이용하여 공리화(axiomatisation)하는 framework를 제안한 논문입니다.

For measurable spaces $(E, \mathcal{E})$ and $(F, \mathcal{F})$, a mapping $K : E \times \mathcal{F} \to [0,1]$ is called a **transition probability kernel** from $(E, \mathcal{E})$ into $(F, \mathcal{F})$ if the mapping $K(x, \cdot)$ is a probability measure on $(F, \mathcal{F})$ for every $x \in E$ and the mapping $K(\cdot, B)$ is measurable for every set $B \in \mathcal{F}$.

Notation

- $(\Omega, \mathcal{H}, \mathbb{P}) = (\times_{t \in T} E_t, \otimes_{t \in T} \mathcal{E}_t, \mathbb{P})$ is a probability space.
- $\mathcal{P}(T)$ is the power set of T.
- For $S \in \mathcal{P}(T)$, $\mathcal{H}_S$ is the sub-σ-algebra of $\mathcal{H} = \otimes_{t \in T} \mathcal{E}_t$ generated by measurable rectangles $\times_{t \in T} A_t$, where $A_t \in \mathcal{E}_t$ differs from E_t only for $t \in S$.
- Ω_S is the subspace $\times_{s \in S} E_s$ of $\Omega = \times_{t \in T} E_t$.
- For $U \subseteq S \subseteq T$, π_{SU} is the natural projection from Ω_S onto Ω_U.

Introduction

- A **causal space** is defined as the quadruple $(\Omega, \mathcal{H}, \mathbb{P}, \mathbb{K})$.
- $(\Omega, \mathcal{H}, \mathbb{P}) = (\times_{t \in T} E_t, \otimes_{t \in T} \mathcal{E}_t, \mathbb{P})$ is a probability space.
- $\mathbb{K} = \{K_S : S \in \mathcal{P}(T)\}$ is the **causal mechanism**: collection of transition probability kernels K_S from $(\Omega, \mathcal{H}_S)$ into $(\Omega, \mathcal{H})$.
- K_S is the **causal kernel** on $\mathcal{H}_S$, which satisfies the following axioms:

  (i) for all $A \in \mathcal{H}$ and $\omega \in \Omega$, $K_\emptyset(\omega, A) = \mathbb{P}(A)$

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

$$K_S(\omega, A \cap B) = 1_A(\omega) K_S(\omega, B)$$

  (iii) for any $A \in \mathcal{H}$, $K_S(\omega, A)$ only depends on the $\pi_{TS}(\omega)$ component of ω.

Decomposition

For $S \in \mathcal{P}(T)$, we denote $\omega_S = \pi_{TS}(\omega)$.

Then note that $\Omega = \Omega_S \times \Omega_{T \setminus S}$ and for any $\omega \in \Omega$, we can decompose it into components as $\omega = \big(\omega_S, \omega_{T \setminus S}\big)$.

Recall that for any $A \in \mathcal{H}$, $K_S(\omega, A) = K_S\left(\left(\omega_S, \omega_{T \setminus S}\right), A\right)$ only depends on the first ω_S component of $\omega = \big(\omega_S, \omega_{T \setminus S}\big)$.

As a slight abuse of notation, we will sometimes write $K_S\left(\omega_S, A\right)$ for conciseness.

Example 1 (Altitude $\to$ temperature)

Example. Let $E_1 = E_2 = \mathbb{R}$, and $\mathcal{E}_1, \mathcal{E}_2$ be Lebesgue σ-algebras on E_1 and E_2. Each $e_1 \in E_1$ and $e_2 \in E_2$ respectively represent the altitude(m) and temperature($^\circ C$) of a random location.

$\mathbb{P}$ is a gaussian measure with mean vector $\begin{pmatrix} 1000 \\ 10 \end{pmatrix}$ and covariance matrix $\begin{pmatrix} 300 & -15 \\ -15 & 10 \end{pmatrix}$.

For each $e_1 \in E_1$ and $A \in \mathcal{E}_2$, we let $K_1(e_1, A)$ be the conditional measure of $\mathbb{P}$ given e_1 with mean $\frac{1200-e_1}{20}$ and variance $\frac{1}{4}$, and $K_2(e_2, B) = \mathbb{P}(B \times E_2)$ for each $B \in \mathcal{E}_1$, i.e. Gaussian measure with mean 1000 and variance 300.

Example 2 (Rice amount $\leftrightarrow$ price)

Using causal model, a situation with "cyclic" causal relationship can be naturally represented.

Example. Consider the relationship between the amount of rice in the market and its price per kg.

$E_1 = E_2 = \mathbb{R}_+$

$\mathcal{E}_1, \mathcal{E}_2$ are Lebesgue σ-algebras and $\mathbb{P}$ is negatively correlated multivariate normal distribution, which implies that the more rice there is in the market, the lower the price without any intervention.

Example 2 (Rice amount $\leftrightarrow$ price)

The causal kernel is $K_1(e_1, x) = \frac{2}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-(e_1+1)}{0.5}\right)^2}$ which implies that if the amount of rice in the market is fixed at e_1, then the price of rice will stabilize with mean $e_1 + 1$ and standard deviation 0.5.

Similarly, $K_2(e_2, y) = \frac{2}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y-e_2}{0.5}\right)^2}$ which implies that if the price of rice is fixed at e_2, then the amount of rice in the market will stabilize with mean e_2 and standard deviation 0.5

Appendix

3.2 Potential Outcomes (PO) Framework

In the PO framework, the treatment and outcome variables of interest are fixed in advance. Although much of the literature begins with individual units, these units are in the end i.i.d. copies of random variables under the stable unit treatment value assumption (SUTVA), and that is how we begin.

Denote by $(\tilde{\Omega}, \tilde{\mathcal{H}}, \tilde{\mathbb{P}})$ the underlying probability space. Let $Z : \tilde{\Omega} \to \mathcal{Z}$ be the "treatment" variable, taking values in a measurable space $(\mathcal{Z}, \mathfrak{Z})$. Then for each value z of the treatment, there is a separate random variable $Y_z : \tilde{\Omega} \to \mathcal{Y}$, called the "potential outcome given $Z = z$" taking values in a measurable space $(\mathcal{Y}, \mathfrak{Y})$; we also have the "observed outcome", which is the potential outcome consistent with the treatment, i.e. $Y = Y_Z$. The researcher is interested in quantities such as the "average treatment effect", $\tilde{\mathbb{E}}[Y_{z_1} - Y_{z_2}]$, where $\tilde{\mathbb{E}}$ is the expectation with respect to $\tilde{\mathbb{P}}$, to measure the causal effect of the treatment. Often, there are other, "pre-treatment variables" or "covariates", which we denote by $X : \tilde{\Omega} \to \mathcal{X}$, taking values in a measurable space $(\mathcal{X}, \mathfrak{X})$. Given these, another object of interest is the "conditional average treatment effect", defined as $\tilde{\mathbb{E}}[Y_{z_1} - Y_{z_2} \mid X]$.

It is relatively straightforward to construct a causal space that can carry this framework. We define $\Omega = \mathcal{Z} \times \mathcal{Y} \times \mathcal{X}$ and $\mathcal{H} = \mathfrak{Z} \otimes \mathfrak{Y} \otimes \mathfrak{X}$. We also define $\mathbb{P}$, for each $A \in \mathfrak{Z}$, $B \in \mathfrak{Y}$ and $C \in \mathfrak{X}$, as $\mathbb{P}(A \times B \times C) = \tilde{\mathbb{P}}(Z \in A, Y \in B, X \in C)$. As for causal kernels, we are essentially only interested in $K_Z(z, B)$ for $B \in \mathfrak{Y}$, and we define these to be $K_Z(z, B) = \tilde{\mathbb{P}}(Y_z \in B)$.