Moving beyond I.I.D.

Causal de Finetti and OOV Generalization

Siyuan Guo

Max Planck Institute for Intelligent Systems
University of Cambridge

April 18th, 2023

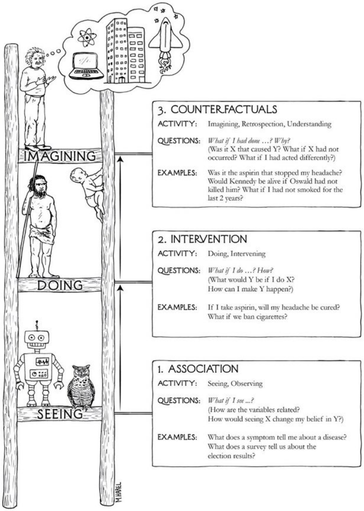


Figure 1: Pearl’s Ladder of Causality

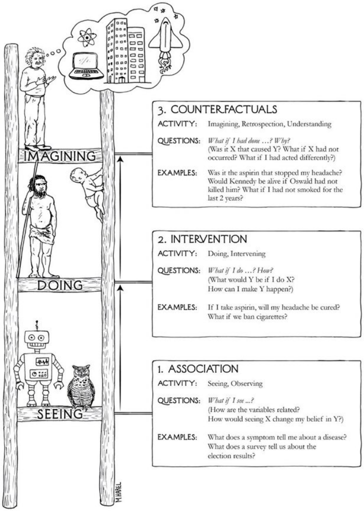
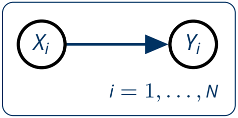


Figure 2: Pearl's Ladder of Causality

I.I.D.



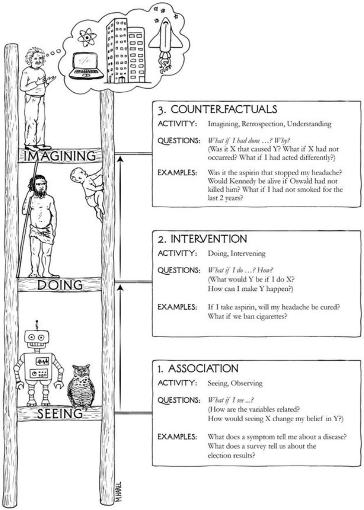
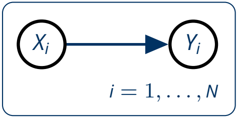


Figure 3: Pearl's Ladder of Causality

I.I.D.



Exchangeable

Let $X_1, X_2, \ldots, X_n$ be a finite sequence of random variables. For any permutation π of $\{1, \ldots, n\}$, it satisfies:

$$\mathbb{P}(X_{\pi(1)}, \ldots, X_{\pi(n)}) = \mathbb{P}(X_1, \ldots, X_n) \quad (1)$$

Then the finite sequence is **exchangeable**. **Infinite exchangeable** sequence if above holds for any $N \in \mathbb{N}$.

De Finetti

Let $(X_n)_{n \in \mathbb{N}}$ be an infinite sequence of binary[1] random variables. The sequence is **exchangeable** *if and only if* there **exists** a latent variable θ such that $X_1, X_2, \dots$ are conditionally i. i. d. given θ.

$$\mathbb{P}(\mathbf{x}_1, \dots, \mathbf{x}_n) = \int \prod_{i=1}^{n} p(\mathbf{x}_i | \theta) d\mu(\theta) \tag{2}$$

[1] De Finetti holds for categorical and continuous random variables.

De Finetti

Let $(X_n)_{n\in\mathbb{N}}$ be an infinite sequence of binary random variables. The sequence is **exchangeable** *if and only if* there **exists** a latent variable θ such that $X_1, X_2, \ldots$ are conditionally i. i. d. given θ.

$$\mathbb{P}(\mathbf{x}_1, \ldots, \mathbf{x}_n) = \int \prod_{i=1}^{n} p(\mathbf{x}_i|\theta) d\mu(\theta) \tag{3}$$

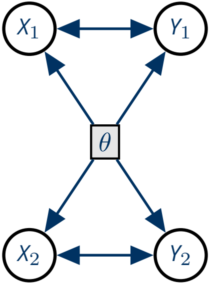
Bayesian Model

- $\mathcal{D} = \{x_1, \ldots, x_n\}$
- Statistical Model $\mathcal{M} = \{P(\cdot \mid \theta) \mid \theta \in \mathcal{T}\}$
- Prior $\theta \sim \pi$

De Finetti Applications

Let $(X_n, Y_n)_{n\in\mathbb{N}}$ be an infinite exchangeable sequence of binary random variables. De Finetti representation theorem states that

$$\mathbb{P}(\mathbf{x}_1, \mathbf{y}_1, \ldots, \mathbf{x}_n, \mathbf{y}_n) = \int \prod_{i=1}^{n} p(\mathbf{x}_i, \mathbf{y}_i|\theta) d\mu(\theta) \tag{4}$$



Independent Causal Mechanism

The causal generative process of a system's variables is composed of autonomous modules that do **not inform** and do **not influence** each other.

$$\mathbb{P}(X_1, \dots, X_n) = \prod_i \underbrace{\mathbb{P}(X_i \mid \mathbf{PA}_i)}_{\text{causal conditional}} \tag{5}$$

Example

$$\implies \quad \text{“}\mathbb{P}(Y \mid X) \perp\!\!\!\perp \mathbb{P}(X)\text{”}$$

Causal De Finetti

Let $\{X_n, Y_n\}_{n\in\mathbb{N}}$[2] be an infinite sequence of binary random variables.
The sequence is

- exchangeable, and
- $\forall n \in \mathbb{N} : Y_{[n]} \perp\!\!\!\perp X_{n+1} \mid X_{[n]}$[3] $\quad \rightarrow \quad$ "P(Y | X) $\perp\!\!\!\perp$ $P(X)$"

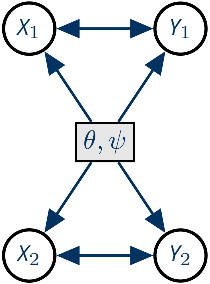
if and only if there exists two independent latent variables θ, ψ such that $X_1, X_2, \ldots$ are conditionally i. i. d. given θ and $Y_1, Y_2, \ldots$ are conditional i. i. d. given ψ and its corresponding X_i.

$$\mathbb{P}(\mathbf{x}_1, \mathbf{y}_1, \ldots, \mathbf{x}_n, \mathbf{y}_n) = \int \prod_{i=1}^{n} p(\mathbf{y}_i \mid \mathbf{x}_i, \psi) p(\mathbf{x}_i \mid \theta) d\mu(\theta) d\nu(\psi) \tag{6}$$

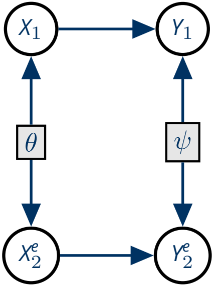
[2] Can extend to multivariate version
[3] $[n] := \{1, \ldots, n\}$

Disentangle the Latents

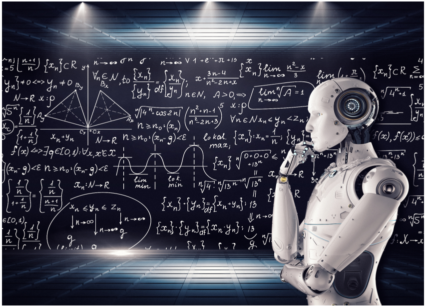
De Finetti:



Causal De Finetti:



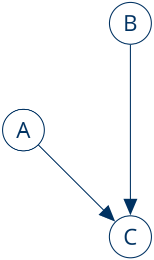
Generalization



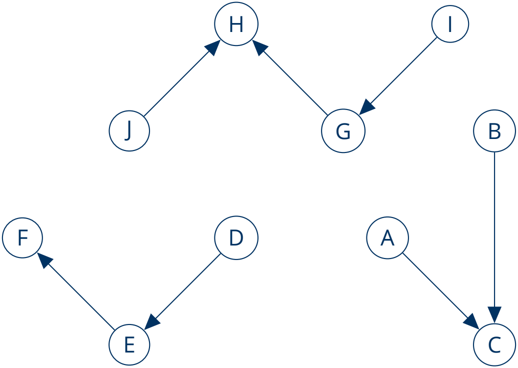
Out-of-Variable Generalization

Out-of-Variable Generalization

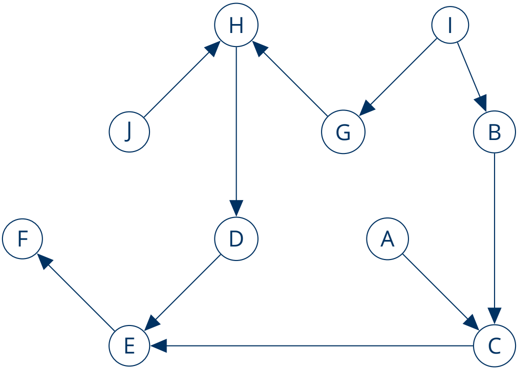
Out-of-Variable Generalization



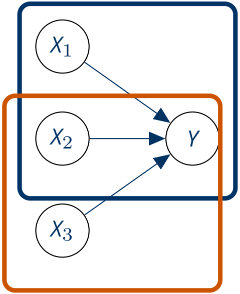
Out-of-Variable Generalization



Out-of-Variable Generalization



Out-of-Variable Genralization



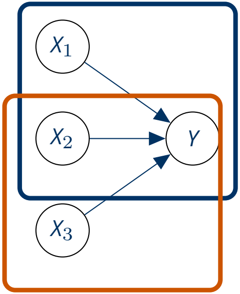
$$\mathcal{D} := (X_1, X_2, Y), (X_2, X_3)$$

$$f_S(x_1, x_2) = \mathbb{E}[Y \mid X_1, X_2]$$

$$f_T(x_2, x_3) = \mathbb{E}[Y \mid X_2, X_3]$$

Figure 4: **Toy example**: (a) the blue box includes variables observed in the source environment, and the orange box those in the target environment. A directed edge represents a causal relationship between two variables. The goal is to improve the zero-shot (i.e., without additional data) prediction of Y in the target environment using the source environment.

Out-of-Variable Generalization



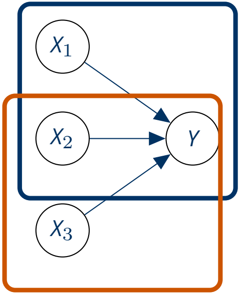
- Consider $Y = \phi(X_1, X_2, X_3) + \epsilon$
- Residual Distribution: $Y - f_S(X_1, X_2)$
- Moments of the residual distribution:

$$\mathbb{E}[(Y - f_S(x_1, x_2))^n \mid x_1, x_2]$$

- Entangled interaction between noise ϵ and $\frac{\partial \phi}{\partial X_3}\Big|_{x_1, x_2, \mu_3}$

$$\mathbb{E}[(Y - f_S(x_1, x_2))^n \mid x_1, x_2] = \sum_{k=0}^{n} \binom{n}{k} \mathbb{E}[\epsilon^k] (\frac{\partial \phi}{\partial X_3}\Big|_{x_1, x_2, \mu_3})^{n-k} \mathbb{E}[(X_3 - \mu_3)^{n-k}]$$

Out-of-Variable Generalization



- Consider $Y = \phi(X_1, X_2, X_3) + \epsilon$
- Residual Distribution: $Y - f_S(X_1, X_2)$
- Moments of the residual distribution:

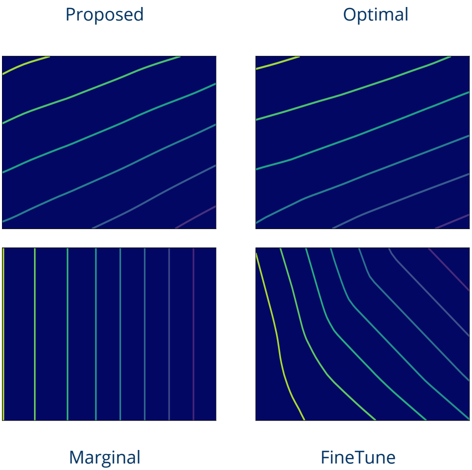
$$\mathbb{E}[(Y - f_S(x_1, x_2))^n \mid x_1, x_2]$$

- Entangled interaction between noise ϵ and $\frac{\partial \phi}{\partial X_3}\Big|_{x_1, x_2, \mu_3}$

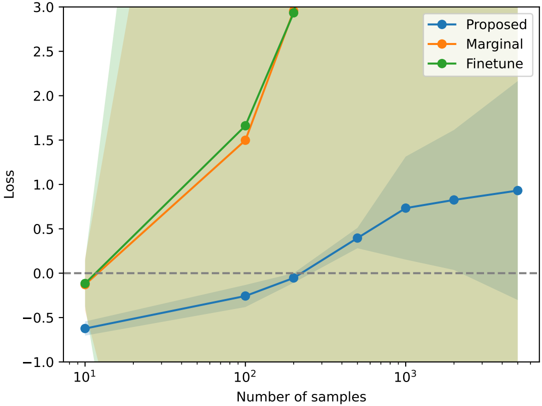
When $n = 3$:

$$\mathbb{E}[(Y - f_S(x_1, x_2))^3 \mid x_1, x_2] = (\frac{\partial \phi}{\partial X_3}\Big|_{x_1, x_2, \mu_3})^3 \mathbb{E}[(X_3 - \mu_3)^3] + \mathbb{E}[\epsilon^3]$$

Experiments



Experiments



Conclusions

Causal de Finetti: On the Identification of Invariant Causal Structure in Exchangeable Data

Siyuan Guo[a,b,1], Viktor Tóth[a,1], Bernhard Schölkopf[b], and Ferenc Huszár[a]

[a]University of Cambridge; [b]Max Planck Institute for Intelligent Systems

---

Out-of-Variable Generalization

Siyuan Guo *†‡ Jonas Wildberger † Bernhard Schölkopf †