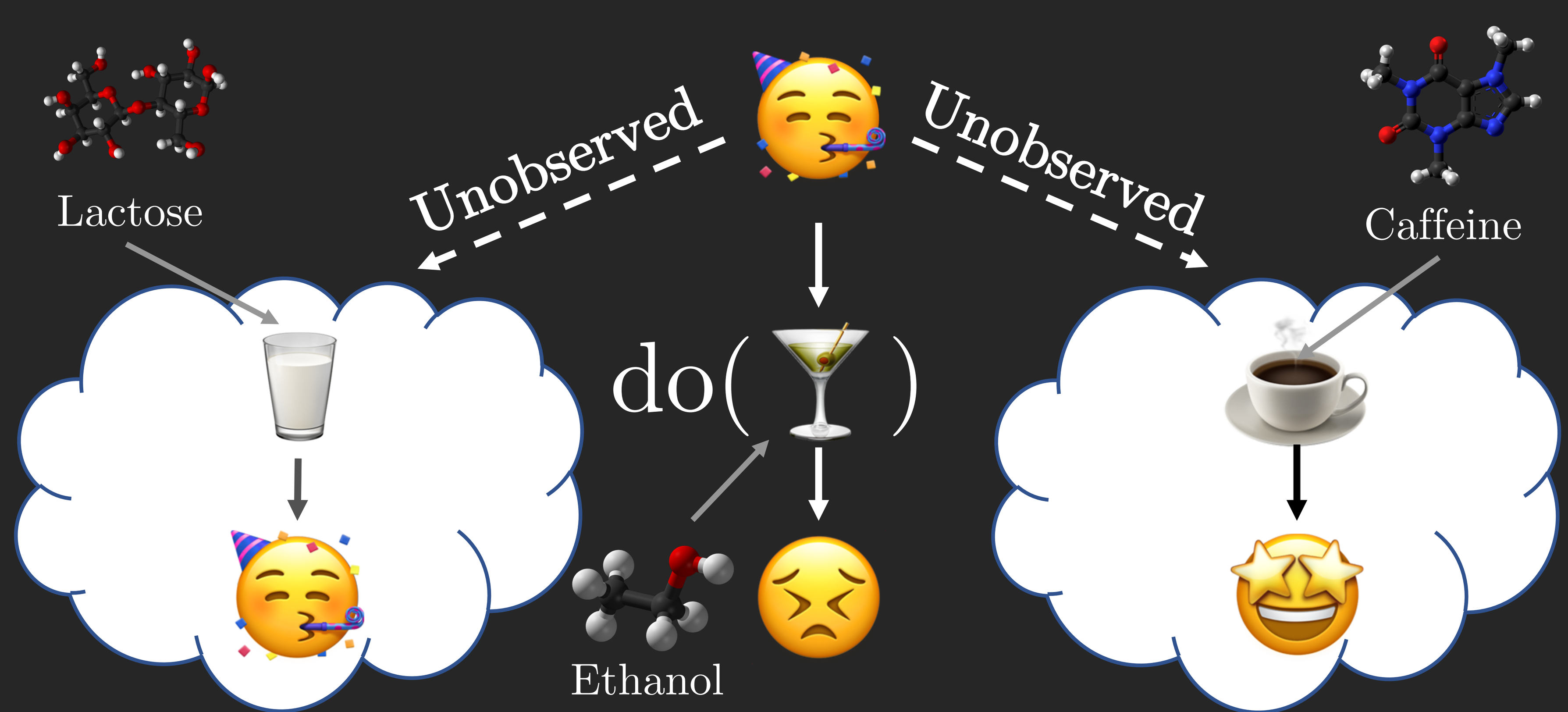
Causal Effect Inference for Structured Treatments

Jean Kaddour, Yuchen Zhu, Qi Liu, Matt Kusner, Ricardo Silva

Why: Motivation

- Imagine you (described by $\mathbf{X}$) had a Martini and felt unwell afterwards
- If you had drunk something else, you might have felt much better
- **Goal**: Estimate the effect of changing the drink $\mathbf{T}$ on the expected wellbeing Y

CATE: τ(☕, 🍸, 🥳) $= \mathbb{E}[Y \mid$ 🥳, do(☕)$] - \mathbb{E}[Y \mid$ 🥳, do(🍸)$]$



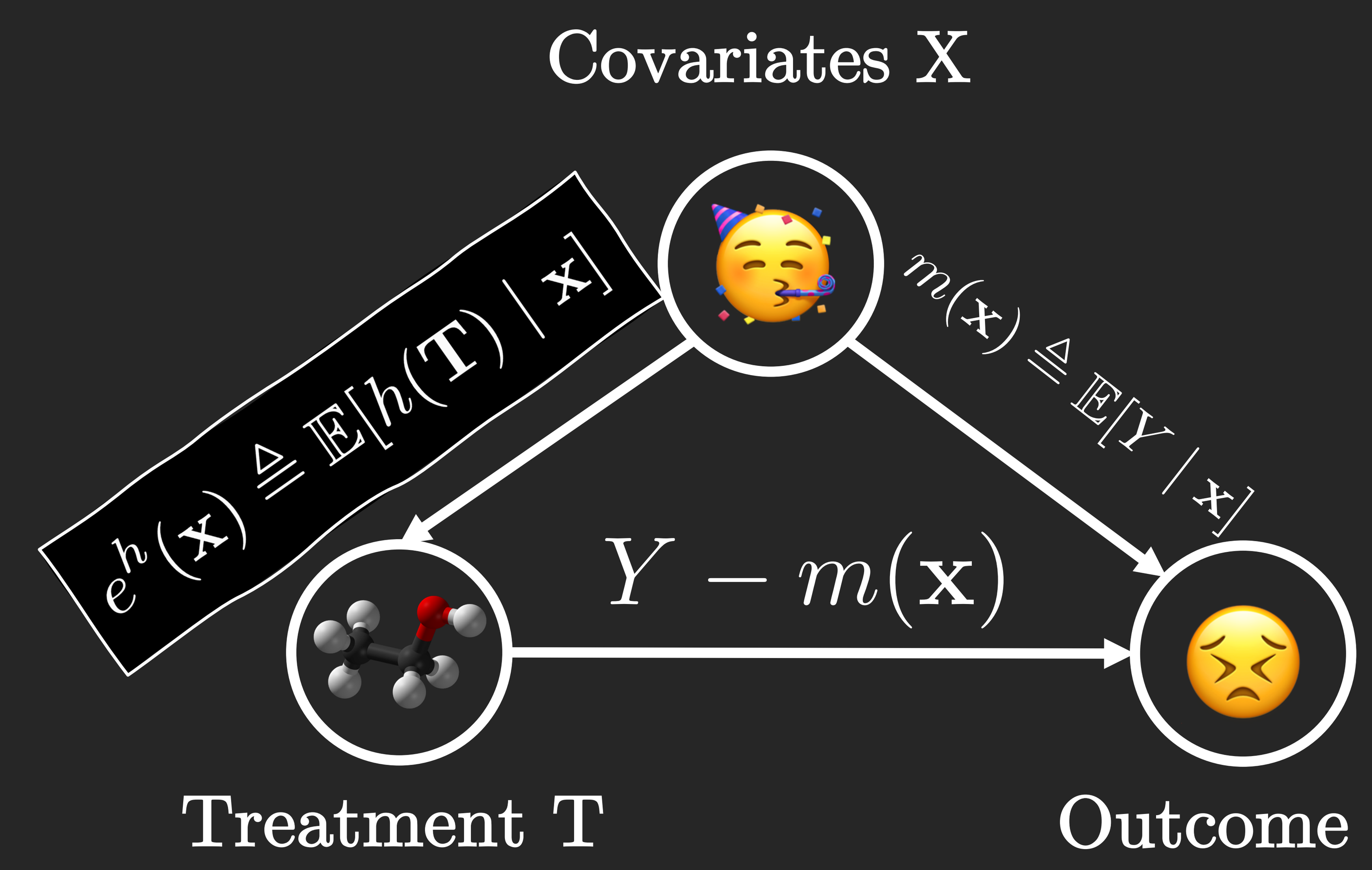
How: Generalized Robinson Decomposition (GRD)

Causal effect as function of treatment features T

$$Y - m(\mathbf{X}) = g(\mathbf{X})^\top \left(h(\mathbf{T}) - e^h(\mathbf{X})\right) + \varepsilon$$

Idea: *Propensity features* $e^h(\mathbf{x}) \triangleq \mathbb{E}[h(\mathbf{T}) \mid \mathbf{x}]$

s.t. mean outcome $m(\mathbf{x}) \triangleq \mathbb{E}[Y \mid \mathbf{x}] = g(\mathbf{x})^\top e^h(\mathbf{x})$



Quasi-Oracle Convergence Guarantee

GRD achieves same error bounds as an oracle who has ground-truth knowledge of both nuisance components $e^h(\mathbf{x})$ and $m(\mathbf{x})$

CATE estimator converges at almost $n^{-1/2}$ rate (fastest rate possible), as long as nuisance functions converge at $n^{-1/4}$ rate.

Assumptions:

- Orthogonal fixed feature maps of covariates $\Psi(\cdot_\mathbf{x})$ and treatments $\Phi(\cdot_\mathbf{t})$
- Overlap on these features $\mathcal{P}_{\Psi(X)\times\Phi(T)} > 0$

Then:

$$\hat{f}(\cdot_\mathbf{x}, \cdot_\mathbf{t}) := \Psi(\cdot_\mathbf{x})^T \Theta \Phi(\cdot_\mathbf{t})$$

$$\downarrow \tilde{O}\left(n^{-\frac{1}{2(1+p)}}\right)$$

$$f^*(\cdot_\mathbf{x}, \cdot_\mathbf{t}) := \mathbb{E}[Y \mid \cdot_\mathbf{x}, \cdot_\mathbf{t}]$$

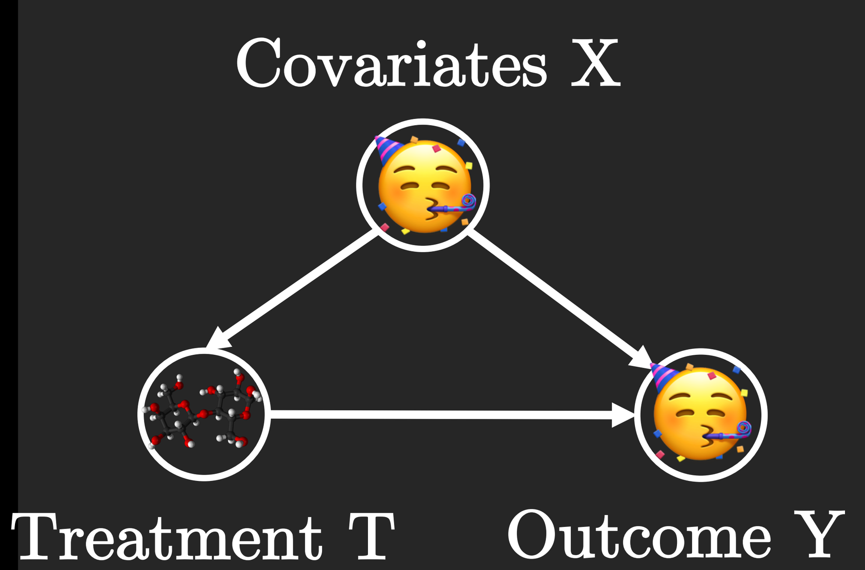
As long as

$$\hat{m}(\cdot_\mathbf{x}) \xrightarrow{O(n^{-1/4})} m(\cdot_\mathbf{x})$$

$$\hat{e}^h(\cdot_\mathbf{x}) \xrightarrow{O(n^{-1/4})} e^h(\cdot_\mathbf{x})$$

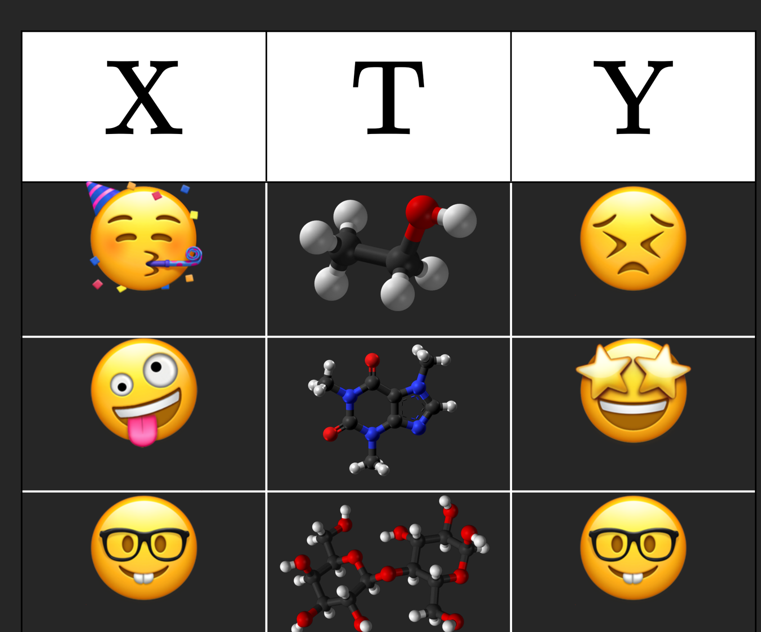
Setup: CATE of Structured Treatments $\mathbf{T} \in \mathcal{T}$

Causal Graph



+

Observational data



+ Assumptions: Overlap, Unconfoundedness

= CATE: τ(⚛, ⚛, 🥳) $= \mathbb{E}[Y \mid$ 🥳, ⚛$] - \mathbb{E}[Y \mid$ 🥳, ⚛$]$

Algorithm: Structured Intervention Networks

Stage 1: Learn parameters of $\widehat{m}_{\boldsymbol{\theta}}(\mathbf{X})$ based on MSE objective

$$J_m(\boldsymbol{\theta}) = \sum_{i=1}^{m} \left(y_i - \widehat{m}_{\boldsymbol{\theta}}(\mathbf{x}_i)\right)^2$$

Stage 2: Alternate between optimizing $\widehat{g}_\psi(\mathbf{X}), \widehat{h}_{\boldsymbol{\phi}}(\mathbf{T})$ and $\hat{e}^h_\eta(\mathbf{X})$

- **a**: Freeze $\widehat{m}_{\boldsymbol{\theta}}(\mathbf{X})$ and $\hat{e}^h_\eta(\mathbf{X})$ to optimize $\widehat{g}_\psi(\mathbf{X}), \widehat{h}_{\boldsymbol{\phi}}(\mathbf{T})$ based on

$$J_{g,h}(\boldsymbol{\phi}, \boldsymbol{\psi}) = \sum_{i=1}^{n} \left(y_i - \left\{\widehat{m}_{\boldsymbol{\theta}}(\mathbf{x}_i) + \widehat{g}_\psi(\mathbf{x}_i)^\top \left(\widehat{h}_\phi(\mathbf{t}_i) - \hat{e}^h_{\boldsymbol{\eta}}(\mathbf{x}_i)\right)\right\}\right)^2$$

- **b**: Freeze $\widehat{m}_{\boldsymbol{\theta}}(\mathbf{X})$ and $\widehat{g}_\psi(\mathbf{X}), \widehat{h}_{\boldsymbol{\phi}}(\mathbf{T})$ to optimize $\hat{e}^h_\eta(\mathbf{X})$ based on

$$J_{e^h}(\boldsymbol{\eta}) = \sum_{i=1}^{n} \sum_{j=1}^{d} \left(\widehat{h}_\phi(\mathbf{t}_i)^{(j)} - \hat{e}^h_{\boldsymbol{\eta}}(\mathbf{x}_i)^{(j)}\right)^2$$

Exemplary Empirical results

Task: Predicting in/out-sample CATEs

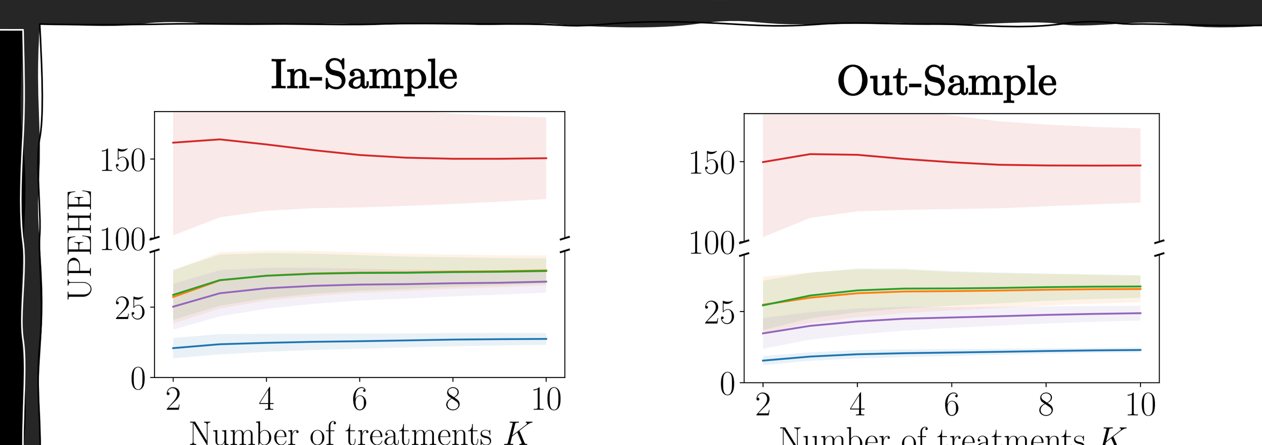
Data: The Cancer Genomic Atlas
X: Gene expression data of cancer patients
T: Molecular graphs from QM9[2] database

Metric: Unweighted expected Precision in Est. of Het. Effects

$$\epsilon_{\text{UPEHE}} \triangleq \int_{\mathcal{X}} \left(\widehat{\tau}(\mathbf{t}', \mathbf{t}, \mathbf{x}) - \tau(\mathbf{t}', \mathbf{t}, \mathbf{x})\right)^2 \mathrm{d}\mathbf{x}$$

Baselines

- GraphITE
- Regression (GNN/CAT)
- Zero Effect



SIN — GraphITE — GNN — CAT — Zero

Code: https://github.com/jeankaddour/SIN