Synthetic Potential Outcomes and Causal Mixture Identifiability

Bijan Mazaheri [1,3] Chandler Squires [2] Caroline Uhler [3,4]

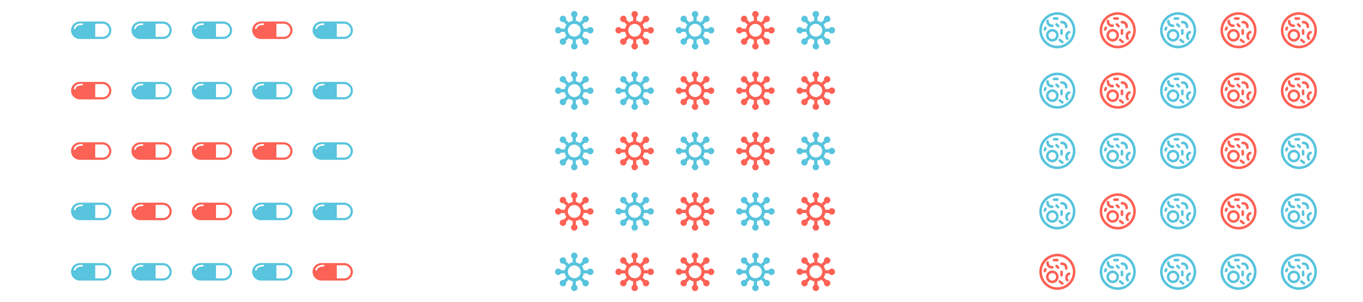
[1]Dartmouth Engineering [2]Carnegie Mellon University [3]Broad Institute of MIT and Harvard [4]MIT

POSTER PAPER TALK

Mixtures of Treatment Effects

Motivation

Heterogeneity in Treatment | Heterogeneity in Subject | Confounded Heterogeneity



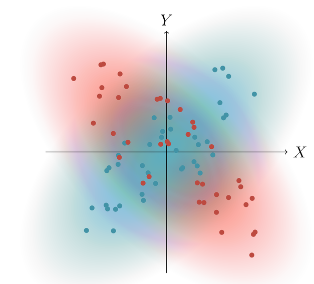
E.g. Quality Control in Treatments | E.g. Vaccine Effectiveness Relative to Viral Strain | E.g. Bacterial Resistance to Prescribed Antibiotic

Main Result

We formulate the **mixture of treatment effects (MTE)** problem for recovering a *latently* heterogeneous causal response and provide the **first known solution**.

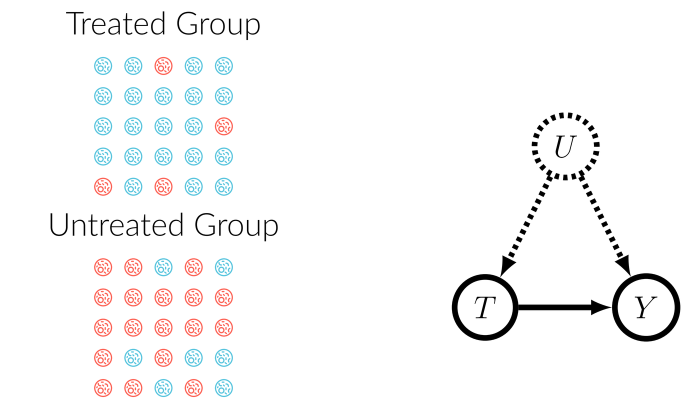
Key Computational Challenges

Distribution Overlap



Algorithms for recovering Heterogeneous Treatment Effects (HTEs), such as causal forests (Wager and Athey, 2018), are analogous to clustering methods and cannot handle distribution overlap.

Mixtures Confound Relationships

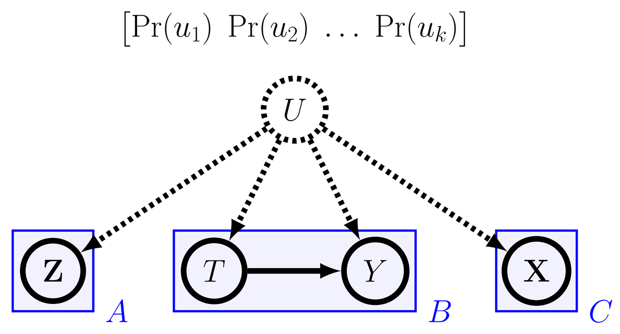


Traditional mixture model algorithms can handle distribution overlap, but do not deconfound causal relationships.

Critical Ideas

Linear Algebra Interpretation

$[\Pr(u_1)\ \Pr(u_2)\ \dots\ \Pr(u_k)]$



$$\begin{bmatrix}\Pr(a_1 \mid u_1)\\ \Pr(a_1 \mid u_2)\\ \vdots \\ \Pr(a_1 \mid u_k)\end{bmatrix} \quad \begin{bmatrix}\Pr(b_1 \mid u_1)\\ \Pr(b_1 \mid u_2)\\ \vdots \\ \Pr(b_1 \mid u_k)\end{bmatrix} \quad \begin{bmatrix}\Pr(c_1 \mid u_1)\\ \Pr(c_1 \mid u_2)\\ \vdots \\ \Pr(c_1 \mid u_k)\end{bmatrix}$$

$$\vdots \qquad \vdots \qquad \vdots$$

$$\begin{bmatrix}\Pr(a_{k_1} \mid u_1)\\ \Pr(a_{k_1} \mid u_2)\\ \vdots \\ \Pr(a_{k_1} \mid u_k)\end{bmatrix} \quad \begin{bmatrix}\Pr(b_{k_2} \mid u_1)\\ \Pr(b_{k_2} \mid u_2)\\ \vdots \\ \Pr(b_{k_2} \mid u_k)\end{bmatrix} \quad \begin{bmatrix}\Pr(c_{k_3} \mid u_1)\\ \Pr(c_{k_3} \mid u_2)\\ \vdots \\ \Pr(c_{k_3} \mid u_k)\end{bmatrix}$$

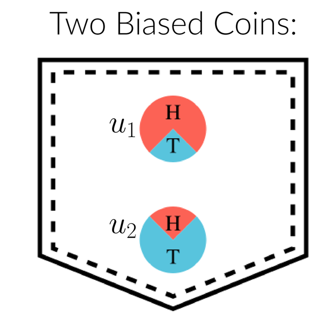
Tensor Decompositions (Allman, Matias, and Rhodes, 2009)

For discrete $A \perp\!\!\!\perp B \perp\!\!\!\perp C \mid U$. with cardinalities $|U| = k$, $|A| = k_1$, $|B| = k_2$, $|C| = k_3$, the parameters $\Pr(a \mid u), \Pr(b \mid u), \Pr(c \mid u), \Pr(u)$ are identifiable using a decomposition into rank 1 tensors if

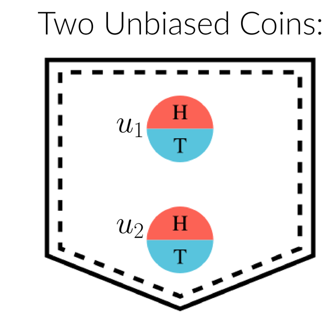
$$k_1 + k_2 + k_3 \geq 2k + 2 \tag{1}$$

Higher Order Moments and Heterogeneity

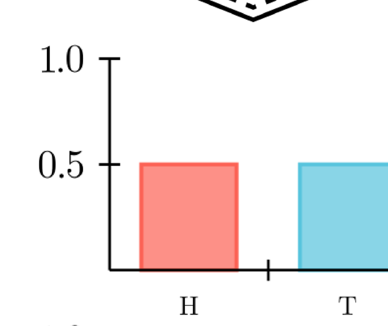
Two Biased Coins:



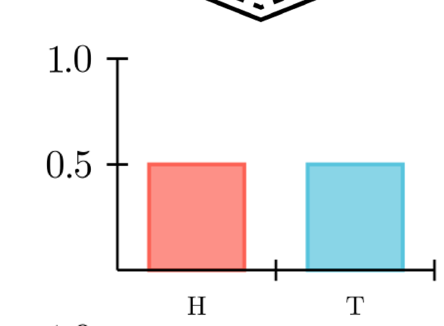
Two Unbiased Coins:



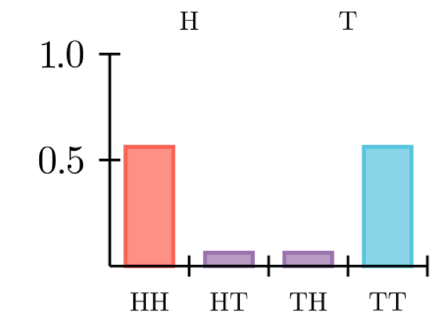
One flip:



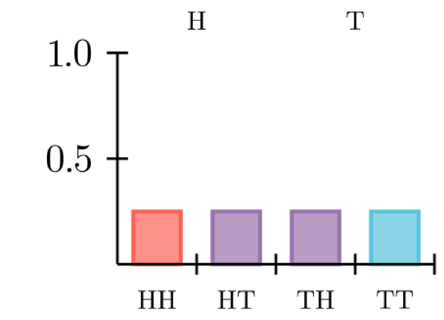
$$\mathbb{E}[R^{\odot 2}] = [\Pr(u_1)\ \Pr(u_2)] \begin{bmatrix}\mathbb{E}[R \mid u_1]\\ \mathbb{E}[R \mid u_2]\end{bmatrix}$$



Two flips:



$$\mathbb{E}[R] = [\Pr(u_1)\ \Pr(u_2)] \begin{bmatrix}\mathbb{E}[R \mid u_1]\\ \mathbb{E}[R \mid u_2]\end{bmatrix}^{\odot 2}$$

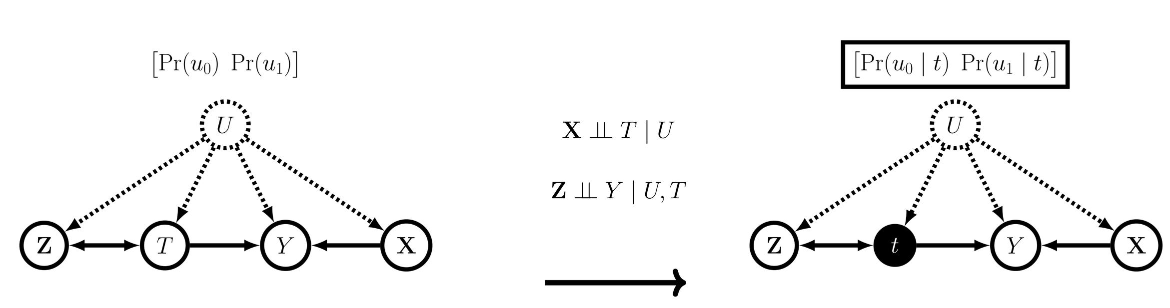


Identifiability of k-Coin Mixtures (Rabani, Schulman, and Swamy, 2014)

$\mathbf{P}[U]$ and distinct $\mathbf{E}[R \mid U]$ can be uniquely identified using $\mathbb{E}(R), \mathbb{E}(R^{\odot 2}), \dots, \mathbb{E}(R^{\odot 2k-1})$.

Synthetic Potential Outcomes

Rederiving Proximal Causal Inference for ATEs



$$\begin{bmatrix}\mathbb{E}(Z_1 \mid u_0)\\ \mathbb{E}(Z_1 \mid u_1)\end{bmatrix} \quad \begin{bmatrix}\mathbb{E}(Y_0 \mid u_0)\\ \mathbb{E}(Y_0 \mid u_1)\end{bmatrix} \quad \begin{bmatrix}\mathbb{E}(X_0 \mid u_0)\\ \mathbb{E}(X_0 \mid u_1)\end{bmatrix} \qquad \boxed{\begin{bmatrix}\mathbb{E}(Z_1 \mid u_0, t)\\ \mathbb{E}(Z_1 \mid u_1, t)\end{bmatrix}} \quad \boxed{\begin{bmatrix}\mathbb{E}(Y_0 \mid u_0, t)\\ \mathbb{E}(Y_0 \mid u_1, t)\end{bmatrix}} \quad \begin{bmatrix}\mathbb{E}(X_0 \mid u_0)\\ \mathbb{E}(X_0 \mid u_1)\end{bmatrix}$$

$$\begin{bmatrix}\mathbb{E}(Z_0 \mid u_0)\\ \mathbb{E}(Z_0 \mid u_1)\end{bmatrix} \quad \begin{bmatrix}\mathbb{E}(Y_1 \mid u_0)\\ \mathbb{E}(Y_1 \mid u_1)\end{bmatrix} \quad \begin{bmatrix}\mathbb{E}(X_1 \mid u_0)\\ \mathbb{E}(X_1 \mid u_1)\end{bmatrix} \qquad \boxed{\begin{bmatrix}\mathbb{E}(Z_0 \mid u_0, t)\\ \mathbb{E}(Z_0 \mid u_1, t)\end{bmatrix}} \quad \boxed{\begin{bmatrix}\mathbb{E}(Y_1 \mid u_0, t)\\ \mathbb{E}(Y_1 \mid u_1, t)\end{bmatrix}} \quad \begin{bmatrix}\mathbb{E}(X_1 \mid u_0)\\ \mathbb{E}(X_1 \mid u_1)\end{bmatrix}$$

We can **reinterpret** a result from Miao, Geng, and Tchetgen Tchetgen, 2018. We will want to find α_0, α_1 that form a linear combination "synthetically samples" the potential outcome.

$$[\Pr(u_0 \mid t)\ \Pr(u_1 \mid t)] \begin{bmatrix}\mathbb{E}(Y \mid u_0, t)\\ \mathbb{E}(Y \mid u_1, t)\end{bmatrix} = [\Pr(u_0 \mid t)\ \Pr(u_1 \mid t)] \left(\alpha_0 \begin{bmatrix}\mathbb{E}(X_0 \mid u_0)\\ \mathbb{E}(X_0 \mid u_1)\end{bmatrix} + \alpha_1 \begin{bmatrix}\mathbb{E}(X_1 \mid u_0)\\ \mathbb{E}(X_1 \mid u_1)\end{bmatrix}\right)$$

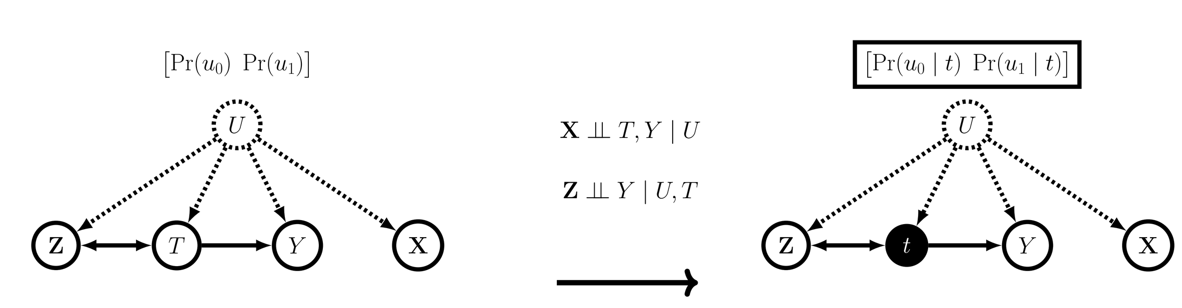
Notice that all of these inner products are observed moments. To get a second equation, we can also do the same thing with the second-order moment, multiplying by Z_0.

$$\mathbb{E}[Y \mid t] = \alpha_0\, \mathbb{E}[X_0 \mid t] + \alpha_1\, \mathbb{E}[X_1 \mid t] \qquad \mathbb{E}[YZ_0 \mid t] = \alpha_0\, \mathbb{E}[X_0 Z_0 \mid t] + \alpha_1\, \mathbb{E}[X_1 Z_0 \mid t]$$

Because $X \perp\!\!\!\perp T \mid U$, we can transfer this linear combination to unconditioned moments to get our desired quantity.

$$\mathbb{E}(Y^{(t)}) = [\Pr(u_0)\ \Pr(u_1)] \begin{bmatrix}\mathbb{E}(Y \mid u_0, t)\\ \mathbb{E}(Y \mid u_1, t)\end{bmatrix} = [\Pr(u_0)\ \Pr(u_1)] \left(\alpha_0 \begin{bmatrix}\mathbb{E}(X_0 \mid u_0)\\ \mathbb{E}(X_0 \mid u_1)\end{bmatrix} + \alpha_1 \begin{bmatrix}\mathbb{E}(X_1 \mid u_0)\\ \mathbb{E}(X_1 \mid u_1)\end{bmatrix}\right) = \alpha_0\, \mathbb{E}[X_0] + \alpha_1\, \mathbb{E}[X_1]$$

Synthetic Potential Outcomes for MTEs



$$\begin{bmatrix}\mathbb{E}(Z_1 \mid u_0)\\ \mathbb{E}(Z_1 \mid u_1)\end{bmatrix} \quad \begin{bmatrix}\mathbb{E}(Y_0 \mid u_0)\\ \mathbb{E}(Y_0 \mid u_1)\end{bmatrix} \quad \begin{bmatrix}\mathbb{E}(X_0 \mid u_0)\\ \mathbb{E}(X_0 \mid u_1)\end{bmatrix} \qquad \boxed{\begin{bmatrix}\mathbb{E}(Z_1 \mid u_0, t)\\ \mathbb{E}(Z_1 \mid u_1, t)\end{bmatrix}} \quad \boxed{\begin{bmatrix}\mathbb{E}(Y_0 \mid u_0, t)\\ \mathbb{E}(Y_0 \mid u_1, t)\end{bmatrix}} \quad \begin{bmatrix}\mathbb{E}(X_0 \mid u_0)\\ \mathbb{E}(X_0 \mid u_1)\end{bmatrix}$$

$$\begin{bmatrix}\mathbb{E}(Z_0 \mid u_0)\\ \mathbb{E}(Z_0 \mid u_1)\end{bmatrix} \quad \begin{bmatrix}\mathbb{E}(Y_1 \mid u_0)\\ \mathbb{E}(Y_1 \mid u_1)\end{bmatrix} \quad \begin{bmatrix}\mathbb{E}(X_1 \mid u_0)\\ \mathbb{E}(X_1 \mid u_1)\end{bmatrix} \qquad \boxed{\begin{bmatrix}\mathbb{E}(Z_0 \mid u_0, t)\\ \mathbb{E}(Z_0 \mid u_1, t)\end{bmatrix}} \quad \boxed{\begin{bmatrix}\mathbb{E}(Y_1 \mid u_0, t)\\ \mathbb{E}(Y_1 \mid u_1, t)\end{bmatrix}} \quad \begin{bmatrix}\mathbb{E}(X_1 \mid u_0)\\ \mathbb{E}(X_1 \mid u_1)\end{bmatrix}$$

We now want to find α_0', α_1' that form a linear combination "synthetically samples" the "second coin flip" of the potential outcome. The process is the same, except that we now copy the product of the synthetic $Y^{(t)}$ on the X's with the true $Y^{(t)}$.

$$[\Pr(u_0 \mid t)\ \Pr(u_1 \mid t)] \begin{bmatrix}\mathbb{E}(Y \mid u_0, t)\\ \mathbb{E}(Y \mid u_1, t)\end{bmatrix}^{\odot 2} = [\Pr(u_0 \mid t)\ \Pr(u_1 \mid t)] \begin{bmatrix}\mathbb{E}(Y \mid u_0, t)\\ \mathbb{E}(Y \mid u_1, t)\end{bmatrix} \odot \left(\alpha_0 \begin{bmatrix}\mathbb{E}(X_0 \mid u_0)\\ \mathbb{E}(X_0 \mid u_1)\end{bmatrix} + \alpha_1 \begin{bmatrix}\mathbb{E}(X_1 \mid u_0)\\ \mathbb{E}(X_1 \mid u_1)\end{bmatrix}\right)$$

Now, we need to use a third-order moment to solve for the coefficients, multiplying by Z_0.

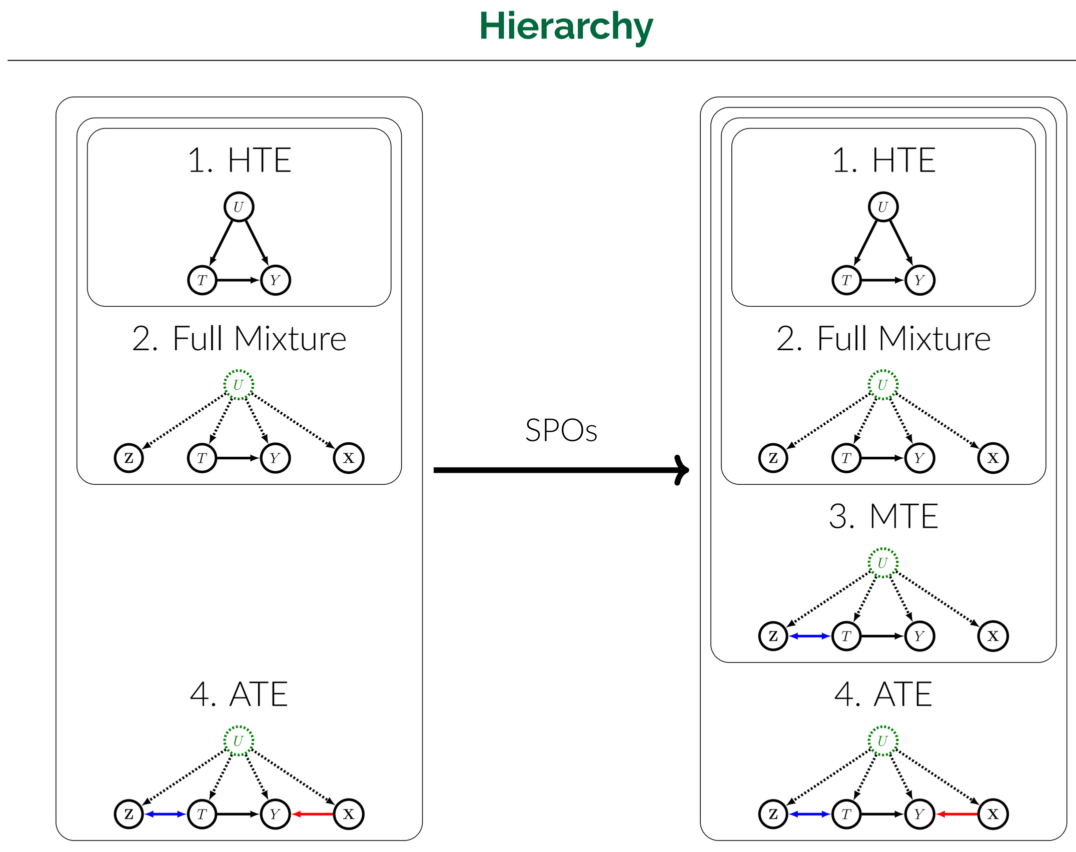
$$\alpha_0\, \mathbb{E}[X_0 Y \mid t] + \alpha_1\, \mathbb{E}[X_1 Y \mid t] = \alpha_0'\, \mathbb{E}[X_0 \mid t] + \alpha_1'\, \mathbb{E}[X_1 \mid t]$$
$$\alpha_0\, \mathbb{E}[X_0 Y Z_0 \mid t] + \alpha_1\, \mathbb{E}[X_1 Y Z_0 \mid t] = \alpha_0'\, \mathbb{E}[X_0 Z_0 \mid t] + \alpha_1'\, \mathbb{E}[X_1 Z_0 \mid t]$$

These coefficients let us compute a higher-order moment of the potential outcome.

$$\mathbb{E}((Y^{(t)})^{\odot 2}) = \alpha_0'\, \mathbb{E}[X_0] + \alpha_1'\, \mathbb{E}[X_1]$$

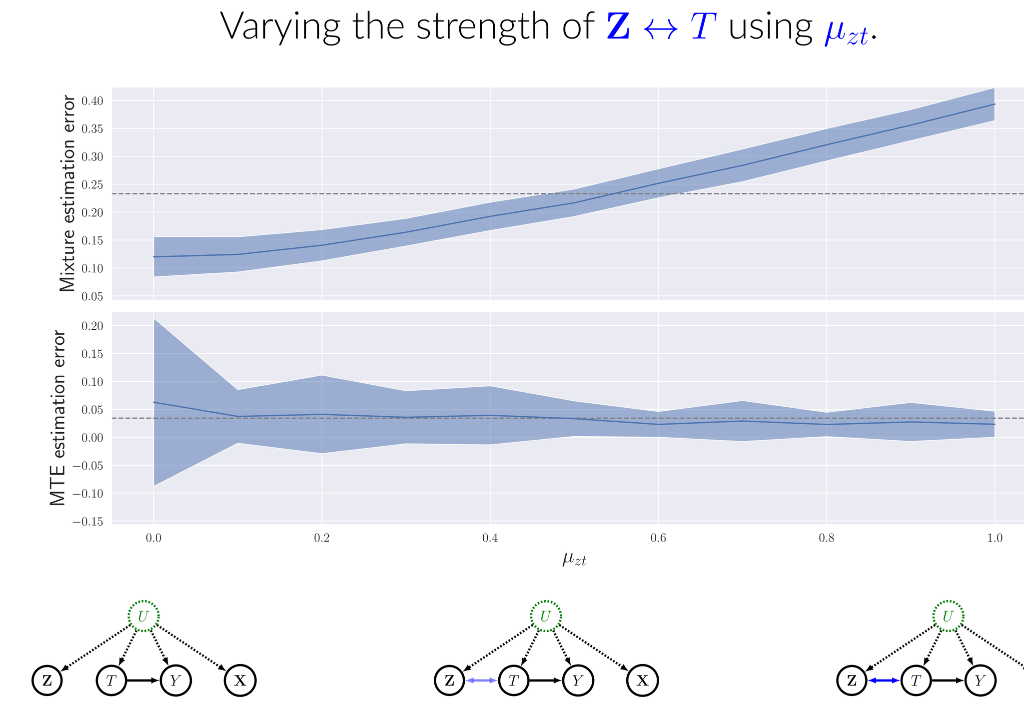
Causal Mixture Identifiability

Hierarchy



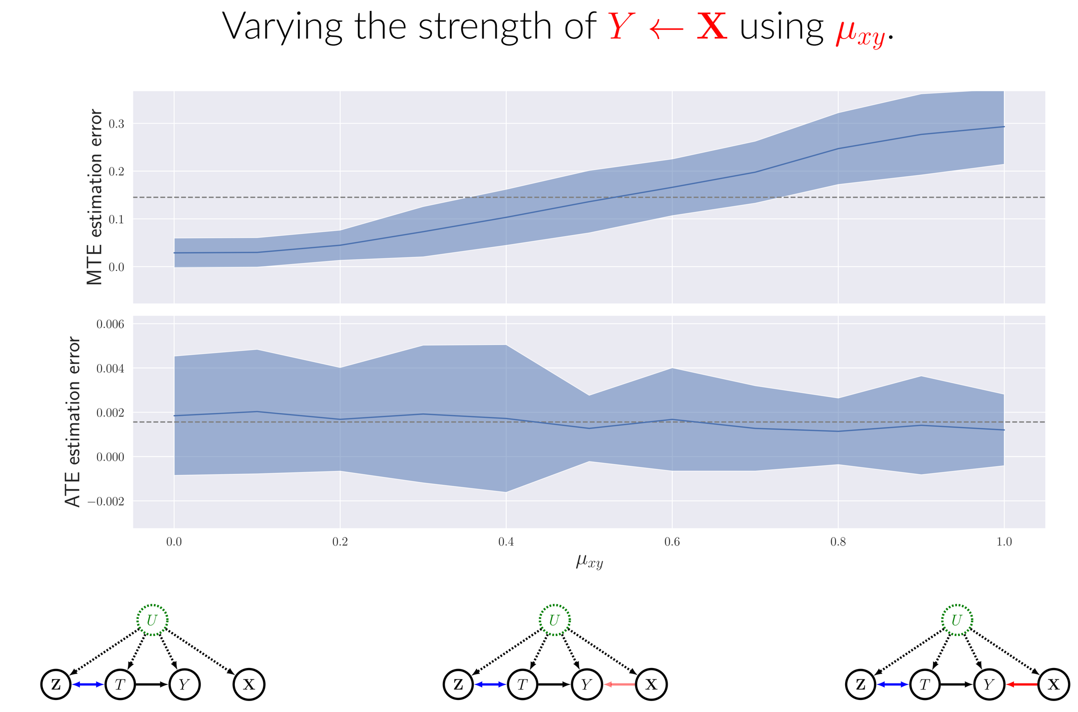
Transition from Level 2 to Level 3

Varying the strength of $\mathbf{Z} \leftrightarrow T$ using μ_{zt}.



Transition from Level 3 to Level 4

Varying the strength of $Y \leftarrow \mathbf{X}$ using μ_{xy}.



References

Allman, Elizabeth S, Catherine Matias, and John A Rhodes (2009). ``Identifiability of parameters in latent structure models with many observed variables''. In.

Miao, Wang, Zhi Geng, and Eric J Tchetgen Tchetgen (2018). ``Identifying causal effects with proxy variables of an unmeasured confounder''. In: *Biometrika* 105.4, pp. 987–993.

Rabani, Yuval, Leonard J Schulman, and Chaitanya Swamy (2014). ``Learning mixtures of arbitrary distributions over large discrete domains''. In: *Proceedings of the 5th conference on Innovations in theoretical computer science*, pp. 207–224.

Wager, Stefan and Susan Athey (2018). ``Estimation and inference of heterogeneous treatment effects using random forests''. In: *Journal of the American Statistical Association* 113.523, pp. 1228–1242.

Thanks

The work for this project was done while Bijan Mazaheri and Chandler Squires were at the Eric and Wendy Schmidt Center at the Broad Institute of MIT and Harvard. Bijan Mazaheri was supported by a postdoctoral fellowship at the Eric and Wendy Schmidt Center. Chandler Squires was partially supported by ONR (N00014-22-1-2116) and DOE-ASCR (DE-SC0023187). Caroline Uhler was partially supported by NCCIH/NIH (1DP2AT012345), ONR (N00014-22-1-2116), DOE-ASCR (DE-SC0023187), the MIT-IBM Watson AI Lab, the Eric and Wendy Schmidt Center at the Broad Institute, and a Simons Investigator Award.