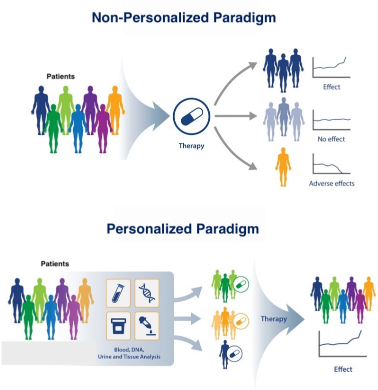
Heterogenous Benefit/Harm Effects from Treatment

Leo Guelman

Motivation for Personalized Decision-Making

- Personalization is founded on the premise that individuals have heterogeneous responses to actions.
- Personalization algorithms aim to improve decision-making by identifying and exploiting this heterogeneity.



- Main ambition of personalized decision-making $\Rightarrow$ Achieve **subject-level** causal effect estimation.
- Primary focus has been on methods that estimate the **Conditional Average Treatment Effects** (CATE).
- CATE estimation can be useful, but does not imply the effects hold at the individual-level.
- **Individual Treatment Effects** (ITEs) are generally non-identifiable, but informative bounds can be obtained on $P(\text{ITE})$.
- **Today's Goal**: Demonstrate how $P(\text{ITE})$ bounds can potentially help improve personalized intervention decisions, relative to using CATE.

Background

Causality has two faces: Necessary and Sufficient

Causality has two faces: Necessary and Sufficient

Probability of Necessity

$$\begin{aligned} PN &\triangleq P(Y_{t'} = \text{false} | T = \text{true}, Y = \text{true}) \\ &\triangleq P(y'_{t'} | t, y) \end{aligned}$$

Causality has two faces: Necessary and Sufficient

Probability of Necessity

$$\begin{aligned} PN &\triangleq P(Y_{t'} = \text{false} | T = \text{true}, Y = \text{true}) \\ &\triangleq P(y'_{t'} | t, y) \end{aligned}$$

Example: A client got a rate discount and renewed her RBC Home Insurance policy. What is the probability she would not have renewed the policy in the absence of the discount?

Probability of Sufficiency

$$PS \triangleq P(y_t | t', y')$$

Probability of Sufficiency

$$PS \triangleq P(y_t | t', y')$$

Example: A client neither got a rate discount nor renewed the insurance policy. What is the probability he would have renewed the policy with the discount?

Probability of Necessity and Sufficiency

$$PNS \triangleq P(y_t, y'_{t'}) \\ \triangleq P(\text{benefit}).$$

Probability of Necessity and Sufficiency

$$\begin{aligned} PNS &\triangleq P(y_t, y'_{t'}) \\ &\triangleq P(\text{benefit}). \end{aligned}$$

- **Example**: What is the probability of a client renewing with rate discount and not renewing without it?

Probability of Necessity and Sufficiency

$$
\begin{aligned}
PNS &\triangleq P(y_t, y'_{t'}) \\
&\triangleq P(\text{benefit}).
\end{aligned}
$$

- **Example**: What is the probability of a client renewing with rate discount and not renewing without it?
- PNS interventions are most relevant in business settings: outcome would not occur without the intervention, and are strong enough to bring the outcome.

PNS relation with CATE and ITE

PNS relation with CATE and ITE

1. PNS and ITE

$$\text{ITE} = y_t - y_{t'}$$

PNS relation with CATE and ITE

1. PNS and ITE

$$\text{ITE} = y_t - y_{t'}$$

$$\text{ITE} = \begin{cases} 1, & P(y_t, y'_{t'}) \\ 0, & P(y_{t'}, y_t) + P(y'_{t'}, y'_t) \\ -1, & P(y_{t'}, y'_t). \end{cases}$$

PNS relation with CATE and ITE

1. PNS and ITE

$$\text{ITE} = y_t - y_{t'}$$

$$\text{ITE} = \begin{cases} 1, & P(y_t, y'_{t'}) \\ 0, & P(y_{t'}, y_t) + P(y'_{t'}, y'_t) \\ -1, & P(y_{t'}, y'_t). \end{cases}$$

Hence, PNS $\triangleq P(\text{benefit}) = P(\text{ITE} > 0)$.

2. PNS and CATE

$$\text{CATE} = P(y_t|X) - P(y_{t'}|X)$$

2. PNS and CATE

$$
\begin{aligned}
\text{CATE} &= P(y_t|X) - P(y_{t'}|X) \\
&= P(y_t, y'_{t'}|x) + P(y_t, y_{t'}|x) - \big(P(y_{t'}, y'_t|x) + P(y_{t'}, y_t|x)\big)
\end{aligned}
$$

2. PNS and CATE

$$
\begin{aligned}
\mathrm{CATE} &= P(y_t|X) - P(y_{t'}|X) \\
&= P(y_t, y'_{t'}|x) + P(y_t, y_{t'}|x) - \big(P(y_{t'}, y'_t|x) + P(y_{t'}, y_t|x)\big) \\
&= \underbrace{P(y_t, y'_{t'}|x)}_{\text{PNS} \triangleq \text{P(benefit|x)}} - \underbrace{P(y_{t'}, y'_t|x)}_{\text{P(harm|x)}}.
\end{aligned}
$$

2. PNS and CATE

$$
\begin{aligned}
\text{CATE} &= P(y_t|X) - P(y_{t'}|X) \\
&= P(y_t, y'_{t'}|x) + P(y_t, y_{t'}|x) - \big(P(y_{t'}, y'_t|x) + P(y_{t'}, y_t|x)\big) \\
&= \quad \underbrace{P(y_t, y'_{t'}|x)}_{\text{PNS} \triangleq \text{P(benefit|x)}} \quad - \underbrace{P(y_{t'}, y'_t|x)}_{\text{P(harm|x)}}.
\end{aligned}
$$

2. PNS and CATE

$$
\begin{aligned}
\text{CATE} &= P(y_t|X) - P(y_{t'}|X) \\
&= P(y_t, y'_{t'}|x) + P(y_t, y_{t'}|x) - \big(P(y_{t'}, y'_t|x) + P(y_{t'}, y_t|x)\big) \\
&= \underbrace{P(y_t, y'_{t'}|x)}_{\text{PNS} \triangleq \text{P(benefit|x)}} - \underbrace{P(y_{t'}, y'_t|x)}_{\text{P(harm|x)}}.
\end{aligned}
$$

Remarks

- P(harm|x) = 0 (Monotonicity) $\Rightarrow$ CATE = P(benefit|x).
- Decomposing CATE into P(benefit|x) and P(harm|x) provides additional information
  - E.g., An intervention may have the same CATE on different client groups, but different P(benefit|x) and P(harm|x) profiles.
- Unlike P(benefit|x), CATE is estimable from experimental data without invoking counterfactual expressions.

Under non-monotonicity, identifying heterogenous benefit/harm effects can be more informative to conventional HTE (CATE) to inform personalized interventions.

Toy example: Home Insurance renewal

- The Home Insurance product team is aiming to send policy renewal e-mail reminders to clients.
- The notification is expected to increase renewal rates through early awareness.

Toy example: Home Insurance renewal

- The Home Insurance product team is aiming to send policy renewal e-mail reminders to clients.
- The notification is expected to increase renewal rates through early awareness.
- They run an A/B Test to assess the validity of this claim:

	Renewed	$\neg$Renewed	Total	Renewal Rate
do(email)	4895	5105	10000	49%
do($\neg$email)	2100	7900	10000	21%
				$\boxed{\text{ATE} = 28\%}$

Table: Average Treatment Effect (ATE)

- ATE reflects the effect on the whole population, which might not hold for different client subgroups.

- ATE reflects the effect on the whole population, which might not hold for different client subgroups.
- CATE estimation (conditioning on X) yields:

		Renewed	¬Renewed	Total	Renewal Rate
X= 1	do(email)	2445	2550	5000	49%
	do(¬email)	1050	3950	5000	21%
					$\boxed{\text{CATE}(X = 1) = 28\%}$
X= 0	do(email)	2450	2550	5000	49%
	do(¬email)	1050	3950	5000	21%
					$\boxed{\text{CATE}(X = 0) = 28\%}$

Table: Conditional Average Treatment Effect (CATE)

- Assume we get access to another sample dataset where clients could sign-up for email renewal notifications.

- Assume we get access to another sample dataset where clients could sign-up for email renewal notifications.
- The observational data shows the following results:

		Renewed	¬Renewed	Total	Renewal Rate
X= 1	email	2000	5000	7000	29%
	¬email	2100	900	3000	70%
X= 0	email	4700	2300	7000	67%
	¬email	2100	900	3000	70%

Table: Renewal Rates by Group

- Combining the experimental and observational data we can conclude:
  - $0.28 \leq P(\text{benefit}|X=1) \leq 0.29$; $P(\text{harm}|X=1) \leq 0.01$.
  - $0.47 \leq P(\text{benefit}|X=0) \leq 0.49$; $0.19 \leq P(\text{harm}|X=0) \leq 0.21$.
- Intuition for non-monotonicity: The renewal notifications triggers an incentive on some $X=0$ clients to shop for better market rates.
- Observational data has value when combined with experimental data. Why?

Derivation of PNS bounds

- In principle, computing counterfactuals such as $PNS = P(y_t, y'_{t'}|x)$, **require** a probabilistic causal model (**PCM**).

- A **PCM** involves:
  - The causal graph (involving endogenous and exogenous variables, V and U, respectively).
  - A parametric specification for the set of functions $\mathrm{F} = \{f_i\}_{i=1}^{n}$, representing $V_i = f_i(\mathrm{pa}_i, U_i)$.
  - A probability distribution defined over $P(U)$.

- "**Require**" $\Rightarrow$ Needed for **identification**.
  - Causal quantity uniquely determined from available data given assumptions.

Factors hindering identification

- **Unobserved Confounding**: Causes and effects influenced by unobserved factors.

- **Sensitivity to** F: The same causal graph can yield different values of counterfactuals depending on F (even for the same $P(U)$).[a]

[a] For instance, see Example 6.19 in Elements of Causal Inference [2017].

PNS in the absence of identification

- We can still obtain bounds on PNS under mild assumptions.
- We will just assume **consistency**:

$$T = t \Rightarrow y = y_t.$$

- A special case of **composition axiom** of counterfactuals:

  The actual world should be closer to itself relative to any world that differs from the actual world.[b]

Remark: We will not be assuming *monotonicity* and/or *unconfoundedness*.

[b] See Galles and Pearl [1998].

PNS bounds

Given *consistency*, our goal is to derive PNS bounds that are:

- **Sharp**: narrowest possible bounds, give this assumption.
- **Symbolic**: closed-form analytic expressions.[c]

[c]Numeric bounds in a general setting are given in Duarte et al. [2021].

PNS bounds: Linear Programming (LP) formulation

- Recall $Y, T \in \{0, 1\}$.
- PCM is not specified, but every PCM induces a distribution on four binary variables: $P(Y, T, Y_t, Y_{t'})$.

PNS bounds: Linear Programming (LP) formulation

- Recall $Y, T \in \{0, 1\}$.
- PCM is not specified, but every PCM induces a distribution on four binary variables: $P(Y, T, Y_t, Y_{t'})$.

$$
\begin{aligned}
\mathrm{PCM}^1 &\Rightarrow P^1(Y, T, Y_t, Y_{t'}) \\
\mathrm{PCM}^2 &\Rightarrow P^2(Y, T, Y_t, Y_{t'}) \\
\mathrm{PCM}^3 &\Rightarrow P^3(Y, T, Y_t, Y_{t'}) \\
&\cdots
\end{aligned}
$$

PNS bounds: Linear Programming (LP) formulation

- Recall $Y, T \in \{0,1\}$.
- PCM is not specified, but every PCM induces a distribution on four binary variables: $P(Y, T, Y_t, Y_{t'})$.

$$\begin{aligned}
\text{PCM}^1 &\Rightarrow P^1(Y, T, Y_t, Y_{t'}) \\
\text{PCM}^2 &\Rightarrow P^2(Y, T, Y_t, Y_{t'}) \\
\text{PCM}^3 &\Rightarrow P^3(Y, T, Y_t, Y_{t'}) \\
&\dots
\end{aligned}$$

- Due to *consistency*, Y is deterministic on the other three variables – e.g., $P(y, t, y_t, y_{t'}) = P(t, y_t, y_{t'}) = P(t, y, y_{t'})$.
- Hence, each distribution is fully specified by $2^3 = 8$ parameters.

Parameters

$$P_{111} = P(y_t, y_{t'}, t)$$
$$P_{110} = P(y_t, y_{t'}, t')$$
$$P_{101} = P(y_t, y'_{t'}, t)$$
$$P_{100} = P(y_t, y'_{t'}, t')$$
$$P_{011} = P(y'_t, y_{t'}, t)$$
$$P_{001} = P(y'_t, y'_{t'}, t)$$
$$P_{000} = P(y'_t, y'_{t'}, t')$$

Parameters

$$
\begin{aligned}
P_{111} &= P(y_t, y_{t'}, t)\\
P_{110} &= P(y_t, y_{t'}, t')\\
P_{101} &= P(y_t, y'_{t'}, t)\\
P_{100} &= P(y_t, y'_{t'}, t')\\
P_{011} &= P(y'_t, y_{t'}, t)\\
P_{001} &= P(y'_t, y'_{t'}, t)\\
P_{000} &= P(y'_t, y'_{t'}, t')
\end{aligned}
$$

Constraints

$$\sum_{i,j,k} P_{ijk} = 1 \tag{1}$$

$$P_{ijk} \geq 0$$

$$P_{111} + P_{101} = P(y, t) \tag{2}$$

$$P_{011} + P_{001} = P(y', t)$$

$$P_{110} + P_{010} = P(y, t')$$

$$p(y_t) = P_{111} + P_{110} + P_{101} + P_{100} \tag{3}$$

$$p(y'_t) = P_{111} + P_{110} + P_{011} + P_{010}$$

Parameters

$$
\begin{aligned}
P_{111} &= P(y_t, y_{t'}, t)\\
P_{110} &= P(y_t, y_{t'}, t')\\
P_{101} &= P(y_t, y'_{t'}, t)\\
P_{100} &= P(y_t, y'_{t'}, t')\\
P_{011} &= P(y'_t, y_{t'}, t)\\
P_{001} &= P(y'_t, y'_{t'}, t)\\
P_{000} &= P(y'_t, y'_{t'}, t')
\end{aligned}
$$

Constraints

$$
\begin{aligned}
\sum_{i,j,k} P_{ijk} &= 1 && (1)\\
P_{ijk} &\geq 0\\
P_{111} + P_{101} &= P(y, t) && (2)\\
P_{011} + P_{001} &= P(y', t)\\
P_{110} + P_{010} &= P(y, t')\\
p(y_t) &= P_{111} + P_{110} + P_{101} + P_{100} && (3)\\
p(y'_t) &= P_{111} + P_{110} + P_{011} + P_{010}
\end{aligned}
$$

- Lower/Upper PNS bounds are obtained by

$$
\begin{aligned}
\text{Min/Max} \quad & PNS = P_{101} + P_{100}\\
\text{s.t.} \quad & (1) - (3)
\end{aligned}
$$

- Solution obtained by applying a vertex enumeration algorithm to the dual LP.

LP solution[d]

$$\max \begin{Bmatrix} 0 \\ P(y_t) - P(y_{t'}) \\ P(y) - P(y_{t'}) \\ P(y_t) - P(y) \end{Bmatrix} \leq PNS \leq \min \begin{Bmatrix} P(y_t) \\ P(y'_{t'}) \\ P(y,t) - P(y',t') \\ P(y_t) - P(y_{t'}) + P(y,t') - P(y,t') \end{Bmatrix}$$

Remarks

- Bounds are sharp since optimization is global.
- The same bounds hold for any subpopulation, by conditioning every term above by covariates X.

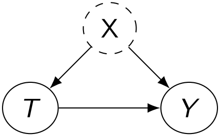
[d] Jin Tian and Judea Pearl [2000].

Finite–Sample PNS bounds[e]

[e] Follows Ang Li, Ruirui Mao, Judea Pearl [2022] with some variants. For details, see https://github.com/leoguelman/pns.

Finite–Sample PNS bounds[e]

Causal Model

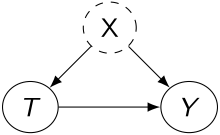


$$
\begin{aligned}
U_{x_i} &\sim \text{Bern}(P_{x_i}), \quad i = \{1, \ldots, 20\}. \\
X_i &:= U_{x_i} \\
T^o &:= \begin{cases} 1 & \text{if } \ \mathrm{X}\beta + U_T > 0.5 \\ 0 & \text{Otherwise.} \end{cases} \\
T_e &\sim \text{Bern}(0.5) \\
Y &:= \begin{cases} 1 & \text{if } \ 0 < \alpha T + \mathrm{X}\gamma + U_Y < 2 \\ 0 & \text{Otherwise.} \end{cases} \\
P_{x_i}, U_T, U_Y &\sim \text{Unif}(0, 1) \\
\beta_i, \gamma_i, \alpha &\sim \text{Unif}(-1, 1)
\end{aligned}
$$

[e]Follows Ang Li, Ruirui Mao, Judea Pearl [2022] with some variants. For details, see https://github.com/leoguelman/pns.

Finite–Sample PNS bounds[e]

Causal Model



$$U_{x_i} \sim \text{Bern}(P_{x_i}), \quad i = \{1, \dots, 20\}.$$

$$X_i := U_{x_i}$$

$$T^o := \begin{cases} 1 & \text{if } \; X\beta + U_T > 0.5 \\ 0 & \text{Otherwise.} \end{cases}$$

$$T_e \sim \text{Bern}(0.5)$$

$$Y := \begin{cases} 1 & \text{if } \; 0 < \alpha T + X\gamma + U_Y < 2 \\ 0 & \text{Otherwise.} \end{cases}$$

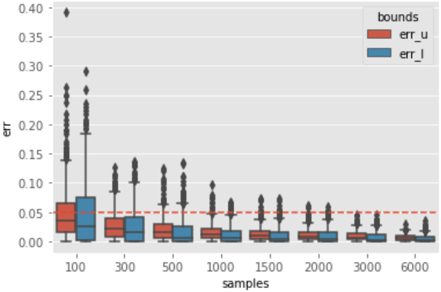
$$P_{x_i}, U_T, U_Y \sim \text{Unif}(0, 1)$$

$$\beta_i, \gamma_i, \alpha \sim \text{Unif}(-1, 1)$$

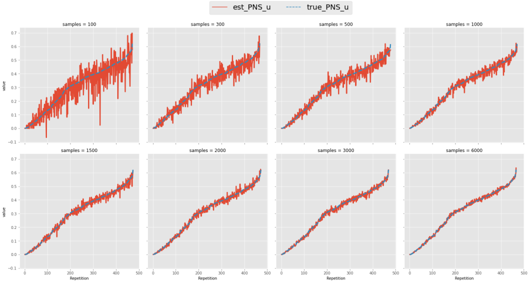
Simulation

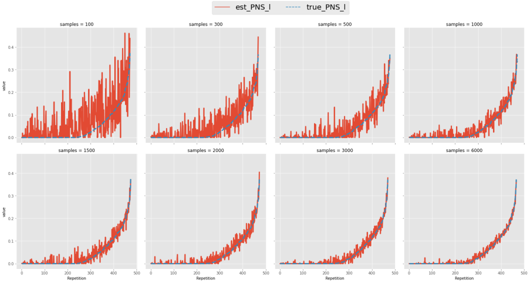
Samples: $N = \{100, 300, 500, 1000, \dots, 6000\}$
Repetitions: $B = 500$
for all $n \in N$ **do**
  for all $b \in B$ **do**
    Draw $(Y_{jb}, X_{jb}, T^e_{jb})$, $j = \{1, \dots, n\}, b = \{1, \dots, B\}$
    Draw $(Y_{jb}, X_{jb}, T^o_{jb})$
    Compute $\hat{PNS}^L_{nb}$, $\hat{PNS}^U_{nb}$, PNS^L_{nb}, PNS^U_{nb}
  end for
  Compute $\mathbf{err}^L_n = \sum_b |\hat{PNS}^L_{nb} - PNS^L_{nb}|/B$,
    $\mathbf{err}^U_n = \sum_b |\hat{PNS}^U_{nb} - PNS^U_{nb}|/B$
end for

[e] Follows Ang Li, Ruirui Mao, Judea Pearl [2022] with some variants. For details, see https://github.com/leoguelman/pns.



- Li, Mao and Pearl [2022] derived sample size requirements to achieve an error rate of at most ϵ, at $1 - \alpha$ confidence-level.
- In the example above, their estimates give $N = 6147$ for $\epsilon = 0.05$ and $\alpha = 0.05$.


est_PNS_u
true_PNS_u
samples = 100
samples = 300
samples = 500
samples = 1000
samples = 1500
samples = 2000
samples = 3000
samples = 6000
value
Repetition


est_PNS_l
true_PNS_l
samples = 100
samples = 300
samples = 500
samples = 1000
samples = 1500
samples = 2000
samples = 3000
samples = 6000
value
Repetition

Work in Progress

- **Goal**: Estimation of heterogenous benefit/harm effects with sharp empirical bounds.
  - Estimation procedure that provides lower error rates on upper/lower bounds, relative to baseline (based on independent estimation of observational and experimental distributions).
  - Effectively deal with non-overlap issues between observational and experimental data.

Appendix

Monotonicity Violations

- A **necessary condition** for monotonicity can be derived by checking that all arguments to the max function in the lower bound of $P(\text{harm}|X)$ are non-positive[f]:

$$\max\begin{Bmatrix} 0 \\ P(y_{t'}|x) - P(y_t|x) \\ P(y|x) - P(y_t|x) \\ P(y_{t'}|x) - P(y|x) \end{Bmatrix} \leq P(\text{harm}|x) \leq \min\begin{Bmatrix} P(y_{t'}|x) \\ P(y'_t|x) \\ P(y', t|x) + P(y, t'|x) \\ P(y_{t'}|x) - P(y_t|x) + P(y, t|x) - P(y', t'|x) \end{Bmatrix}$$

- From the above, we get

$$\text{Monotonicity} \Rightarrow P(y_t|x) \geq P(y|x) \geq P(y_{t'}|x) \tag{4}$$

- By contrapositive, failure to satisfy the consequent of 4 implies monotonicity violations.
- Given $P(Y, T|X)$, what are the hypothetical values of $P(y_t|x)$ and $P(y_{t'}|x)$ that entail violations to the inequality in 4?

[f] $P(\text{harm}|X)$ bounds can be obtained from PNS and CATE using the relation derived in Slide 9.

- Interventional distributions are compatible with observational distributions under the following conditions[g]:

$$
\begin{aligned}
P(y,t|x) \leq P(y_t|x) \leq 1 - P(y',t|x) \\
P(y,t'|x) \leq P(y_{t'}|x) \leq 1 - P(y',t'|x).
\end{aligned}
\tag{5}
$$

Define the **compatible set** as $\mathcal{C} = \Big\{\big(P(y_t|x), P(y_{t'}|x)\big) : \text{Ineq.}(5)\Big\}$.

- Example: Suppose we have the following observational and experimental data
  - $P(t|x) = 0.5$, $P(y|t,x) = 0.5$, $P(y|t',x) = 0.5$.
  - $P(y_t|x) = 0.5$, $P(y_{t'}|x) = 0.5$.

Then, compatibility implies:

$$
\begin{aligned}
0.25 \leq P(y_t|x) \leq 0.75 \\
0.25 \leq P(y_{t'}|x) \leq 0.75.
\end{aligned}
$$

[g] See Eq. 22 in Jin Tian and Judea Pearl [2000].

- Hence, incorporating the necessary condition for monotonicity (Ineq. 4), we get that interventional distributions must satisfy:

$$
\begin{aligned}
0.50 &\leq P(y_t|x) \leq 0.75 \\
0.25 &\leq P(y_{t'}|x) \leq 0.50.
\end{aligned}
$$

- Define the **necessary set** as $\mathcal{N} = \{(P(y_t|x), P(y_{t'}|x)) : \text{Ineq.}(4) \wedge \text{Ineq.}(5)\}$.
- Finally, define the **violation set** as $\mathcal{V} = \mathcal{C} \setminus \mathcal{N}$ – i.e., $\mathcal{V}$ is composed of feasible values of the interventional distributions that would entail violations to monotonicity.