NC STATE UNIVERSITY

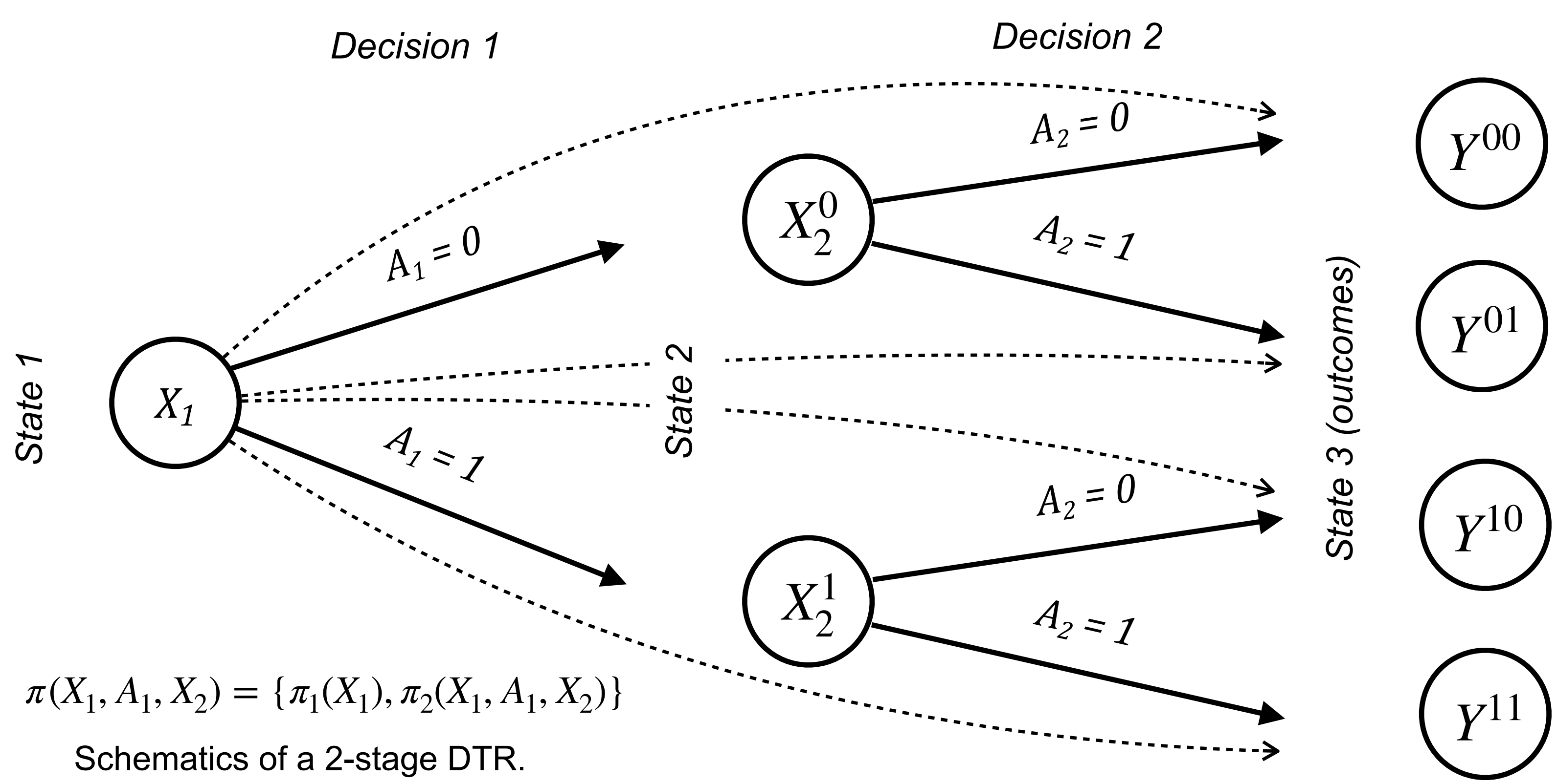
Evaluating and Learning Optimal Dynamic Treatment Regimes under Truncation by Death

Sihyung Park Wenbin Lu Shu Yang
North Carolina State University

The 39th Annual Conference on Neural Information Processing Systems

Dynamic treatment regime (DTR)

- A sequence of decision rules for choosing effective treatments for individual patients.

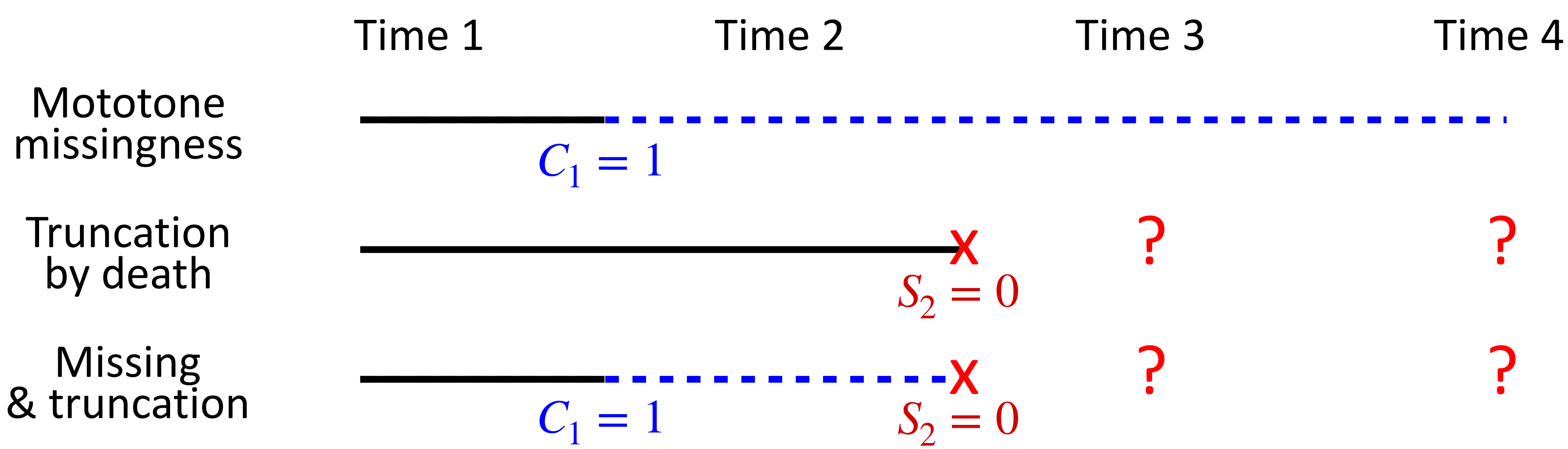


Schematics of a 2-stage DTR.

- While methods for estimating optimal DTRs have proliferated, their applicability is often limited by the incomplete data.

Missingness and Truncation by Death

- **Motonone missingness**: The outcome could have been measured if the subject had not dropped out.
- **Truncation by death**: The outcome cannot be measured because the subject is no longer alive; the outcomes are not simply missing but ill-defined.



Framework

Symbol	Description
$\pi = (\pi_1, \pi_2)$	A deterministic 2-stage binary policy.
X_k	Baseline or intermediate covariates.
A_k	Treatment indicator at k (1 if treated).
C_k	Censorship indicator at k (1 if censored).
S_k	Survival indicator at k (1 if survived).
Y	Outcome of interest.
H_k	History up to k. i.e., $\{\bar{X}_k, \bar{A}_{k-1}\}$.
U	Principal stratum. i.e., $S_2^{00} S_2^{01} S_2^{10} S_2^{10}$.
$p_k^{\bar{a}_k}(\bar{x}_k)$	Survival probability at k. $\mathbb{P}(S_k = 1 \mid \bar{x}_k, \bar{a}_k, \bar{C}_k = \mathbf{0}_k, \bar{S}_{k-1} = \mathbf{1}_{k-1})$.
$\varphi_k^{\bar{a}_k}(\bar{x}_k)$	Joint propensity-censoring probability at k defined as $\mathbb{P}(A_k = a_k, C_k = 0 \mid \bar{x}_k, \bar{a}_{k-1}, \bar{C}_{k-1} = \mathbf{0}_{k-1}, \bar{S}_{k-1} = \mathbf{1}_{k-1})$.
$Q_{Y,2}(\bar{X})$	Q-function at $k = 2$. $\mathbb{E}[Y \mid \bar{X}, \pi(\bar{X}), \bar{C} = (0,0), \bar{S} = (1,1)]$.
$Q_{Y,1}(X_1)$	Q-function at $k = 1$. $\mathbb{E}[Q_{Y,2}(\bar{X}) \mid X_1, \pi_1(X_1), C_1 = 0, S_1 = 1]$.
$Q_{S,2}(\bar{X})$	Eventual survival probability at $k = 2$. $S_1 p_2^{00}(\bar{X})$
$Q_{S,1}(X_1)$	Eventual survival probability at $k = 1$. $p_1^0(X_1)\mathbb{E}[p_2^{00}(\bar{X}) \mid X_1, a_1, C_1 = 0, S_1 = 1]$

Objective: maximize $V_{\mathrm{AS}}(\pi) = E[Y^{\pi} \mid U = 1111]$.

Framework

Assumption 1 (Causal consistency). $Z_1^{A_1} = Z_1$ and $Z_2^{A_1 A_2} = Z_2$ for variables $Z_1 \in \{C_1, S_1, X_2\}, Z_2 \in \{C_2, S_2, Y\}$.

Assumption 2 (Sequential randomization). For all *k=1,2*,

(i) $(S_1^{a_1}, S_2^{a_1 a_2}, C_1^{a_1}, C_2^{a_1 a_2}) \perp\!\!\!\perp A_k \mid H_k$

(ii) $Y^{a_1 a_2} \perp\!\!\!\perp A_k \mid H_k, S_2^{a_1 a_2} = 1$, and $X_2^{a_1} \perp\!\!\!\perp A_k \mid H_k, S_1^{a_1} = 1$

Standard in DTR and causal inference literature

Assumption 3 (Monotonicity).

(i) $S_2^{a_1 a_2} \geq S_1^{a_1}$ and $C_2^{a_1 a_2} \leq C_1^{a_1}$.

(ii) $S_1^1 \geq S_1^0$ and $S_2^{11} \geq S_2^{10}, S_2^{01} \geq S_2^{00}$.

Standard in principal stratification literature

Assumption 4 (Principal ignorability). For $u_1, u_2, u_3 \in \{0,1\}$,

(i) $E[Y^{01} \mid \bar{X}^0, U = 1111] = E[Y^{01} \mid \bar{X}^0, U = u_1 1 u_3 1]$,

$E[Y^{10} \mid \bar{X}^1, U = 1111] = E[Y^{10} \mid \bar{X}^1, U = u_1 u_2 11]$, and

$E[Y^{11} \mid \bar{X}^1, U = 1111] = E[Y^{11} \mid \bar{X}^1, U = u_1 u_2 u_3 1]$.

(ii) $E[g(X_2^1) \mid X_1, U = 1111] = E[g(X_2^1) \mid X_1, U = u_1 u_2 u_3 1]$ and

$E[g(X_2^0) \mid X_1, U = 1111] = E[g(X_2^0) \mid X_1, U = u_1 1 u_3 1]$ for integrable g.

Additionally, positive probabilities, bounded outcome mean, and regularity conditions including convergence rate of nuisance functions are required.

Off-policy evaluation

Under the assumptions,

$$V_{\mathrm{AS}}(\pi) = E[w(X_1)Q_{Y,1}(X_1)] \qquad \text{(Principal Q-learning)}$$

$$= E\left[\frac{\mathbf{1}\{\bar{A} = \pi(\bar{X})\}(1 - C_1)(1 - C_2)S_1S_2}{\varphi_1^\pi(X_1)\varphi_2^\pi(\bar{X})p_1^\pi(X_1)p_2^\pi(\bar{X})}w(X_1)Y\right] \qquad \text{(Principal IPW)}$$

where $w(X_1) = Q_{S,1}(X_1)/E[Q_{S,1}(X_1)]$ is the principal score.

- "Principal" because it targets the mean outcome in the latent always-survivor group.
- e.g. For principal Q-learning, we can train backward-recursively:

$$\hat{Q}_2(\bar{x};\pi) = \hat{E}\left[Y \mid \bar{x}, \pi(\bar{x}), \bar{C} = \mathbf{0}_2, \bar{S} = \mathbf{1}_2\right], \quad \tilde{V}^\pi := \hat{Q}_2(\bar{x};\pi),$$

$$\hat{Q}_1(x_1;\pi) = \hat{E}\left[\tilde{V}^\pi \mid x_1, \pi_1(x_1), C_1 = 0, S_1 = 1\right].$$

Multiply robust evaluation and Learning

Define

$$\phi_D(O) = \frac{\mathbf{1}\{\bar{A} = \bar{C} = \mathbf{0}_2\}}{\varphi_1^0(X_1)\varphi_2^{\mathbf{0}_2}(\bar{X})} S_1 S_2 + \sum_{k=1}^{2} \left\{ \prod_{j=0}^{k-1} \frac{\mathbf{1}\{A_j = C_j = 0\}}{\varphi_j^{\mathbf{0}_j}(\bar{X}_j)} - \prod_{j=0}^{k} \frac{\mathbf{1}\{A_j = C_j = 0\}}{\varphi_j^{\mathbf{0}_j}(\bar{X}_j)} \right\} Q_{S,k},$$

$$\phi_{N(\pi)}(O) = Q_{Y,1}\phi_D(O) + \sum_{k=1}^{2} \prod_{j=1}^{k} \frac{\mathbf{1}\{\pi_j(\bar{X}_j) = A_j\}(1 - C_j)S_j}{\varphi_j^{\pi}(\bar{X}_j) p_j^{\pi}(\bar{X}_j)} \left(Q_{Y,j+1} - Q_{Y,j} \right) Q_{S,1}.$$

Then $\hat{V}_{\mathrm{MR}}(\pi) = \hat{\phi}_{N(\pi)}(O)/\hat{\phi}_D(O)$ is a consistent estimator of $V_{\mathrm{AS}}(\pi)$ if the nuisance model specification satisfies any of the following scenarios:

	$\hat{\varphi}_1$	$\hat{p}_1$	$\hat{\varphi}_2$	$\hat{Q}_{S,2}$	$\hat{Q}_{S,1}$	$\hat{Q}_{Y,2}$	$\hat{Q}_{Y,1}$
(i)	X	X	X	X	X		
(ii)	X		X			X	X
(iii)	X	X		X	X	X	
(iv)	X			X		X	X
(v)		X		X	X	X	X

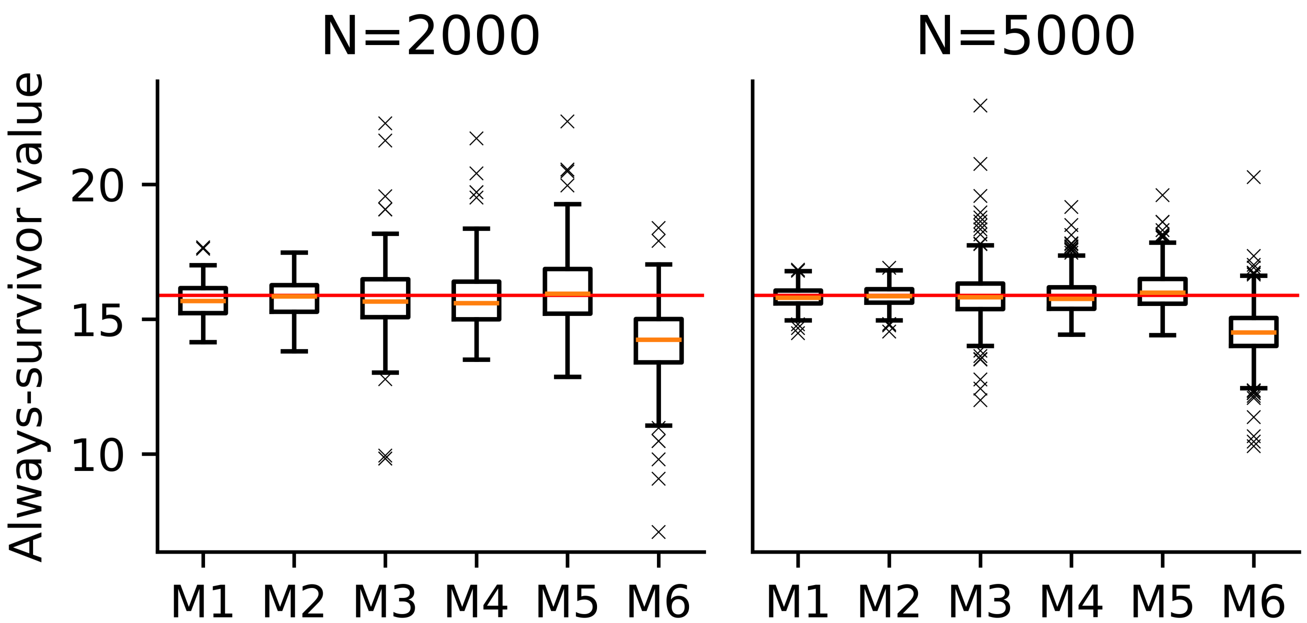
'**X**' indicates correct specification.

If all nuisance models are correct, $\hat{V}_{\mathrm{MR}}(\pi)$ is semiparametric locally efficient.

$n^{1/2}\left(\hat{V}_{\mathrm{MR}}(\hat{\pi}) - V_{\mathrm{AS}}(\pi^*)\right) \xrightarrow{d} \mathcal{N}\left(0, \Upsilon(\pi^*)\right)$ if the policy class Π is not complex*, where $\hat{\pi}$ is the policy maximizing $\hat{V}_{\mathrm{MR}}$, and π^* is the true optimal policy,

Simulation study

We tested on data of size *n=2000* and *n=5000,* six model specification settings with M1-M5 expected to result in unbiased result.



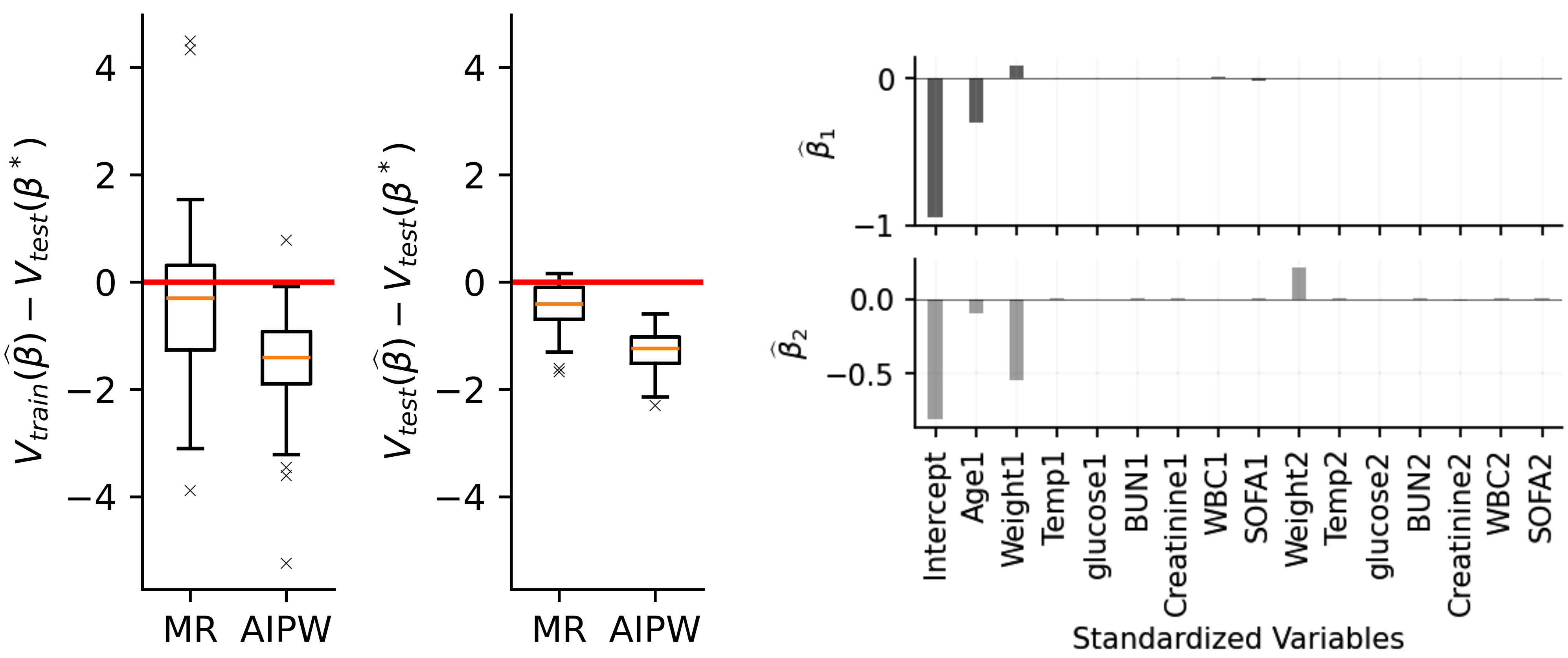
		M1	M2	M3	M4	M5	M6
n=2000	Bias	-0.20	-0.10	-0.15	-0.13	0.22	-1.74
	S.E.	0.63	0.63	1.19	1.11	1.26	1.27
n=5000	Bias	-0.07	-0.02	-0.04	-0.06	0.18	-1.39
	S.E.	0.37	0.37	1.05	0.66	0.74	1.04

		Principal Q	Principal IPW	MR
n=2000	Bias	-0.02	0.13	0.11
	S.E.	0.37	0.86	0.63
n=5000	Bias	-0.07	-0.03	0.06
	S.E.	0.28	0.71	0.48

(Top, Bottom Left) The MR estimator is consistent in M1-M5 settings as expected, with M1 resulting in the most efficient. **(Bottom Right)** The MR estimator is more efficient than the principal IPW estimator under M1.

Application to MIMIC-III

Intervention with mechanical ventilation at each time point was represented by A_k. A class of linear policies with 15 covariates were explored. The MR estimator is compared with the standard augmented IPW estimator.



(Left) The MR estimator is closer to the validation value function. **(Right)** Learned policy is in alignment with previously reported findings, with age and weight being the most important factors. The policy tends to offer mechanical ventilation only in emergent cases.

Discussion

In practice, always-survivors are unknown.

we recommend a constraint optimization, or equivalently,

$$\tilde{\pi} = \underset{\pi \in \Pi}{\text{argmax}}\, \hat{V}_{\mathsf{MR}}(\pi) + \lambda(1 - \hat{S}_t^*(\pi))$$

where $S_t^*(\pi)$ is the expected time-t mortaility rate following policy π.

$\tilde{\pi}$ balances between survival rate and the optimal always-survivor policy.

References

J. Chu, S. Yang, W. Lu (2023). "Multiply robust off-policy evaluation and learning under truncation by death". In: Proceedings of the 40th International Conference on Machine Learning, 202:6195–6227.

G. W. Imbens, J. D. Angrist (1994). "Identification and estimation of local average treatment effects". In: Econometrica, 62(2):467–475.

R. Jiang, W. Lu, R. Song, M. Davidian (2017). "On estimation of optimal treatment regimes for maximizing t-year survival probability". In: Journal of the Royal Statistical Society Series B: Statistical Methodology, 79(4):1165–1185.

Z. Jiang, S. Yang, P. Ding (2022). "Multiply robust estimation of causal effects under principal ignorability". In: Journal of the Royal Statistical Society SeriesB: Statistical Methodology, 84(4):1423–1445.

A. E. Johnson, T. J. Pollard, L. Shen, L. H. Lehman, M. Feng, M. Ghassemi, B. Moody, P. Szolovits, L. Anthony Celi, R. G. Mark (2016). "MIMIC-III, a freely accessible critical care database". In: Scientific data, 3(1): 1–9.

M. Komorowski, L. A. Celi, O. Badawi, A. C. Gordon, A. A. Faisal (2018). "The artificial intelligence clinician learns optimal treatment strategies for sepsis in intensive care". In: Nature medicine, 24(11):1716–1720.