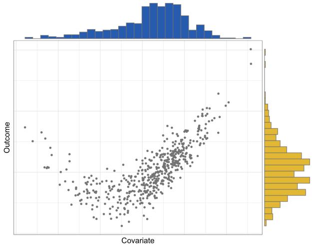
Transportability Index: Inverse Probability Weighting with Neural Network

Xinran Miao and Jiwei Zhao

December 2, 2022

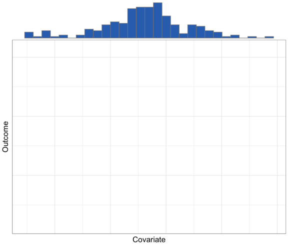
Motivation: model transportability

Source



$$\begin{array}{cccccc} R & 1 & 1 & 1 & \cdots & 1 \\ \mathbf{X} & \mathbf{x}_1 & \mathbf{x}_2 & \mathbf{x}_3 & \cdots & \mathbf{x}_n \\ Y & y_1 & y_2 & y_3 & \cdots & y_n \end{array}$$

Target



$$\begin{array}{ccccc} 0 & 0 & 0 & \cdots & 0 \\ \mathbf{x}_{n+1} & \mathbf{x}_{n+2} & \mathbf{x}_{n+3} & \cdots & \mathbf{x}_{n+m} \\ y_{n+1} & y_{n+2} & y_{n+3} & \cdots & y_{n+m} \end{array}$$

Motivation: model transportability

Notations

- We use subscripts s and t for distributions conditioned on $R = 1$ and $R = 0$, respectively.
- $\mathbb{E}_s(\cdot) = \mathbb{E}(\cdot \mid R = 1)$.
- $\mathbb{E}_t(\cdot) = \mathbb{E}(\cdot \mid R = 0)$.

Problem of interest

- Consider $\boldsymbol{\beta}$ in $\mathbb{E}_t\{\boldsymbol{\xi}(Y, \mathbf{X}; \boldsymbol{\beta})\} = \mathbf{0}$ for some $\boldsymbol{\xi}$.
- For example, $\beta = \mathbb{E}_t(Y)$.

Assumption: covariate shift

Distribution exchangeability

To identify β, a common assumption is

$$p_t(y \mid \mathbf{x}) = p_s(y \mid \mathbf{x}) \tag{1}$$

Identifiability

Then we can identify β as

$$\begin{aligned}\beta &= \mathbb{E}_t(Y) \\ &= \mathbb{E}_t\{\mathbb{E}_t(Y \mid \mathbf{X})\} \\ &= \mathbb{E}_t\{\mathbb{E}_s(Y \mid \mathbf{X})\}\end{aligned}$$

Sensitivity Analysis

Sensitivity assumption

- Since (1) is untestable, sensitivity analysis studies how inference may change when (1) is violated.
- Following [RRS00, SNK+21], we assume

$$p_t(y \mid \mathbf{x}) \propto \rho(y, \mathbf{x}, \gamma) p_s(y \mid \mathbf{x}) \tag{2}$$

with a known sensitivity parameter γ.

- This work considers $\rho(y, \mathbf{x}, \gamma) = \exp(\gamma y)$.
- When $\gamma = 0$, (2) reduces to (1).
- Under (2), β can be viewed as a function of γ.

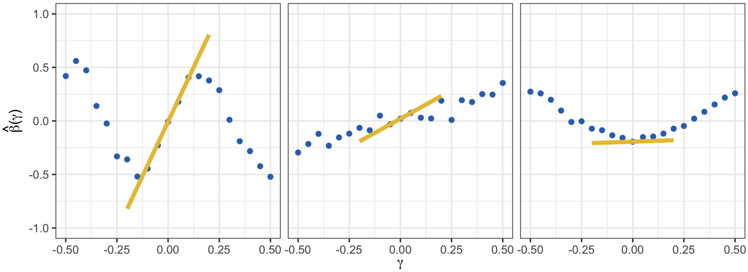
$$\beta(\gamma) = \mathbb{E}_t(Y) = \mathbb{E}_t\left[\frac{\mathbb{E}_s\{\exp(\gamma Y)\, Y \mid \mathbf{X}\}}{\mathbb{E}_s\{\exp(\gamma Y) \mid \mathbf{X}\}}\right]$$

Sensitivity to perturbation

First order Taylor approximation

For γ near zero,

$$\beta(\gamma) \approx \beta(0) + \gamma \cdot \boxed{\left.\frac{\partial \beta(\gamma)}{\partial \gamma}\right|_{\gamma=0}}$$



Transportability index

Definition

We define the transportability index parameter

$$\lambda = \left.\frac{\partial \beta(\gamma)}{\partial \gamma}\right|_{\gamma=0} \tag{3}$$

Form

Under choices of $\rho(y, \mathbf{x}, \gamma) = \exp(\gamma y)$ and $\beta = \mathbb{E}_t(Y)$, λ has the form

$$\lambda = \mathbb{E}_t\{\mathrm{var}_s(Y \mid \mathbf{X})\}.$$

Problem revisit

Goal: Efficiently estimate $\beta(\gamma)$ and λ.

Outline

IPW estimators

Inverse Probability Weighting estimators

$$\widehat{\beta}_{\text{ipw}}(\gamma) = \frac{1}{n}\sum_{i=1}^{n} \frac{\exp(\gamma y_i)\, y_i}{\mathbb{E}_s\{\exp(\gamma Y) \mid \mathbf{x}_i\} w(\mathbf{x}_i)} \tag{4}$$

$$\widehat{\lambda}_{\text{ipw}} = \frac{1}{n}\sum_{i=1}^{n} \frac{\{y_i - \mathbb{E}_s(Y \mid \mathbf{x}_i)\}^2}{w(\mathbf{x}_i)}, \tag{5}$$

where $w(\mathbf{x}) = p_s(\mathbf{x})/p_t(\mathbf{x})$ is the density ratio.

Nuisances

Each parameter includes two nuisances

- We need $w(\mathbf{x})$ and $\mathbb{E}_s(\exp(\gamma Y) \mid x)$ to estimate (4).
- We need $w(\mathbf{x})$ and $\mathbb{E}_s(Y \mid \mathbf{x})$ to estimate (5).

Estimating the density ratio

Estimating $w(\mathbf{x})$ is essentially estimating a conditional probability.

$$\begin{aligned}\mathrm{pr}(R = 1 \mid \mathbf{x}) &= \frac{\pi p_s(\mathbf{x})}{\pi p_s(\mathbf{x}) + (1-\pi)p_t(\mathbf{x})} \\ &= \frac{1}{1 + \frac{1-\pi}{\pi}\frac{1}{w(\mathbf{x})}},\end{aligned}$$

where $\pi = \mathrm{pr}(R = 1)$ can be plugged in as $n/(n+m)$, the proportion of source data.

Estimating nuisances via neural network

When the number of covariates gets large

- In nuisance estimation, parametric models can induce huge bias and traditional non-parametric estimators (e.g., kernel based methods) may collapse as the number of covariates, d, increases.
- Recent studies suggests that, outcome-regression typed estimators are efficient when nuisances are estimated with Artificial Neural Networks (ANN) [CLMZ20, FLM21].
- Artifical Neural Network, also known as Neural Network, extracts linear combinations of the input as features and models the output as their nonlinear transformations [CLMZ20].

Artificial Neural Network (ANN)

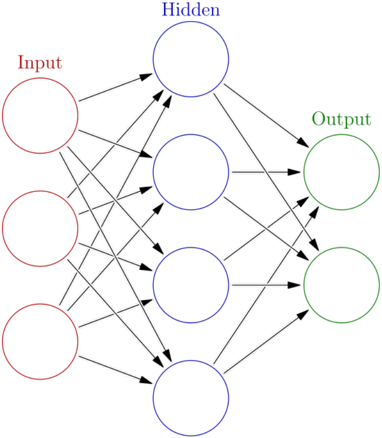


Figure: Illustration of neural network (from Wikipedia).

Semi-parametric Efficiency Lower Bound

IPW esimated by NN

- When nuisances are estimated by neural networks, (4) and (5) achieves semi-parametric efficiency lower bound.
- Efficient influence functions are given by (6) and (7).

$$
\begin{aligned}
\phi_{\text{eff},\boldsymbol{\beta}} =& \frac{r}{\pi} \cdot \frac{\mathbb{E}_s\{\exp(\gamma Y) \mid \mathbf{x}\}\rho(y,\gamma)y - \mathbb{E}_s\{\exp(\gamma Y)\, Y \mid \mathbf{x}\}\exp(\gamma Y)}{w(\mathbf{x})\left[\mathbb{E}_s\{\exp(\gamma Y) \mid \mathbf{x}\}\right]^2} + \\
& \frac{1-r}{1-\pi}\left[\frac{\mathbb{E}_s\{\exp(\gamma Y)\, Y \mid \mathbf{x}\}}{\mathbb{E}_s\{\exp(\gamma Y) \mid \mathbf{x}\}} - \beta_0(\gamma)\right]
\end{aligned}
\tag{6}
$$

$$
\phi_{\text{eff},\lambda} = \frac{r}{\pi w(\mathbf{x})}\left[\{y - \mathbb{E}_s(Y \mid \mathbf{x})\}^2 - \text{var}(Y \mid \mathbf{x})\right] + \frac{1-r}{1-\pi}\{\text{var}_s(Y \mid \mathbf{x}) - \lambda_0\},
\tag{7}
$$

where $\pi = \text{pr}(R=1)$ and $\beta_0(\gamma)$ and λ_0 are true values.

Semi-parametric Efficiency Lower Bound

Remark

Efficient influence functions of both $\beta(\gamma)$ and λ include three nuisances.

1. *Nuisances for $\beta(\gamma)$*
   - $w(\mathbf{x})$,
   - $\mathbb{E}_s\{\exp(\gamma Y) \mid \mathbf{x}\}$, *and*
   - $\mathbb{E}_s\{\exp(\gamma Y)\, Y \mid \mathbf{x}\}$.
2. *Nuisances for λ:*
   - $w(\mathbf{x})$,
   - $\mathbb{E}_s\{Y \mid \mathbf{x}\}$, *and*
   - $var_s(Y \mid \mathbf{x})$.

Outline

Transportability across critical care units

Medical Information Mart for Intensive Care (MIMIC-III) database

MIMIC-III contains hospital admission information of 38,597 distinct adult patients admitted to five types of critical care units between 2001 and 2012.

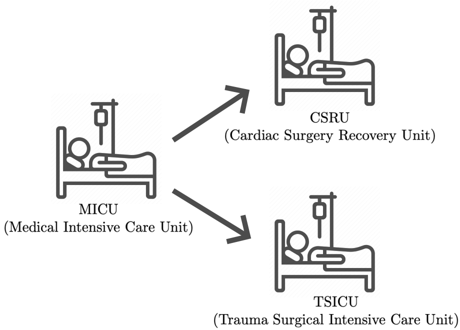


Figure: Source - target split on MIMIC-III data. The icon comes from iconfinder.

Notation

Transfer setting

- We define the source / target split to be the initial admission care unit.
- Source: MICU (Medical Intensive Care Unit).
- Target: CSRU (Cardiac Surgery Recovery Unit) and TSICU (Trauma Surgical Intensive Care Unit).

Outcome Y

- SOFA score: Sequential Organ Failure Assessment.
- SAPS II score: Siplified Acute Physiology Score.

Covariate $\mathbf{X}$

- 14 variables including demographics, chart events and laboratory tests.

Covariate description

Type	Name	Description
Demographics	age	Age of a patient
	gender	Gender of a patient
Chart events	diasbp	Diastolic blood pressure
	glucose	Blood glucose
	resprate	Respiratory rate per minute
	sysbp	Systolic blood pressure
	temp	Body temperature
	hr	Heart rate per minute
Laboratory Tests	hemotocrit	Hematocrit level
	platelets	Platelets count
	redbloodcell	Red blood cell count
	whitebloodcell	Red blood cell count
	urea_n	Blood urea nitrogen
	calcium	Calcium level in blood

Table: Covariate description in MIMIC III data. Chart event and laboratory test variables have been taken average on.

Data structure

Pre-processing

- We only keep the initial admission record for each patient to guarantee independence.
- Abnormal values in covariates are removed (e.g., age greater than 200).
- Missing values are imputed by their average (proportion < 0.05).
- Both covariates and outcomes are standardized prior to estimation.

Sample size

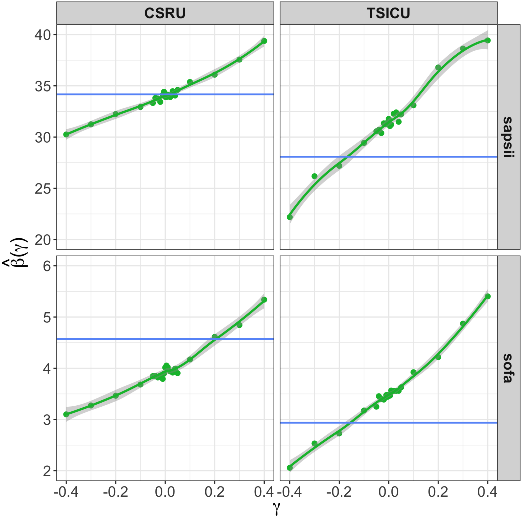
- Source: 14,825 patients.
- Target: 7,856 and 4,727 patients from CSRU and TSICU, respectively.

Implementation details

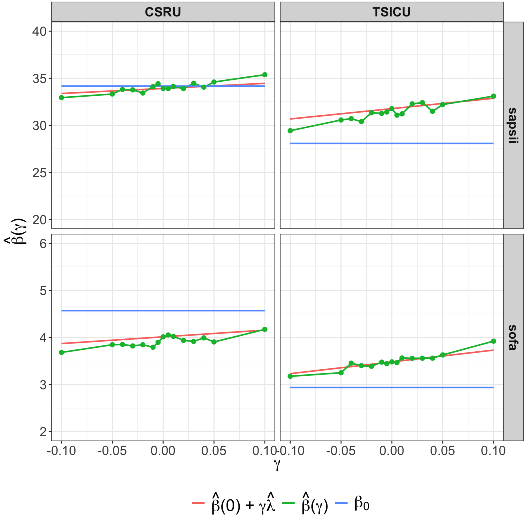
Estimation with neural networks

- All nuisances are estimated by one-layer neural networks with ReLU activations.
- Hyperparameters are chosen via a grid search.
- Hyperparameters under consideration: number of hidden units, number of epochs, batch size, learning rate.

Result : $\beta = \mathbb{E}_t(Y)$



Result : transportability index



Outline

Discussion

Takeaways

- We propose a general framework to assess the covariate shift assumption in a transfer setting.
- The proposed measure, transportability index, calibrates the sensitivity to change in quantity of interest with respect to a small perturbation.
- Our estimation procedure resembles inverse-probability-weighting estimate and chooses to estimate nuisances with neural networks. The estimation achieves semi-parametric efficiency lower bound.

Limitations and Future Work

Generalizing the form of parameters

- The form of λ replies on $\rho(y, \mathbf{x}, \gamma) = \exp(\gamma y)$ and the simplicity of target parameter $\beta = \mathbb{E}_t(Y)$.
- We only considered first-order Taylor expansion. One can similarly define sensitivity parameters for higher order terms.

Estimation

- Instead of impowering the estimation of nuisances (e.g., with NN), one may deal with the curse of dimensionality by dimensionality reduction or assuming a sparsity in signals.

A broader context

- The same idea of sensitivity can be applied in causal inference to assess the validation / violation of the ignorability assumption.
- It would be interesting to consider multiple sources.

The End

Thank you!

References I

[CLMZ20] Xiaohong Chen, Ying Liu, Shujie Ma, and Zheng Zhang, *Efficient estimation of general treatment effects using neural networks with a diverging number of confounders*, arXiv preprint arXiv:2009.07055 (2020).

[FLM21] Max H Farrell, Tengyuan Liang, and Sanjog Misra, *Deep neural networks for estimation and inference*, Econometrica **89** (2021), no. 1, 181–213.

[RRS00] James M Robins, Andrea Rotnitzky, and Daniel O Scharfstein, *Sensitivity analysis for selection bias and unmeasured confounding in missing data and causal inference models*, Statistical models in epidemiology, the environment, and clinical trials, Springer, 2000, pp. 1–94.

[SNK+21] Daniel O Scharfstein, Razieh Nabi, Edward H Kennedy, Ming-Yueh Huang, Matteo Bonvini, and Marcela Smid, *Semiparametric sensitivity analysis: Unmeasured confounding in observational studies*, arXiv preprint arXiv:2104.08300 (2021).