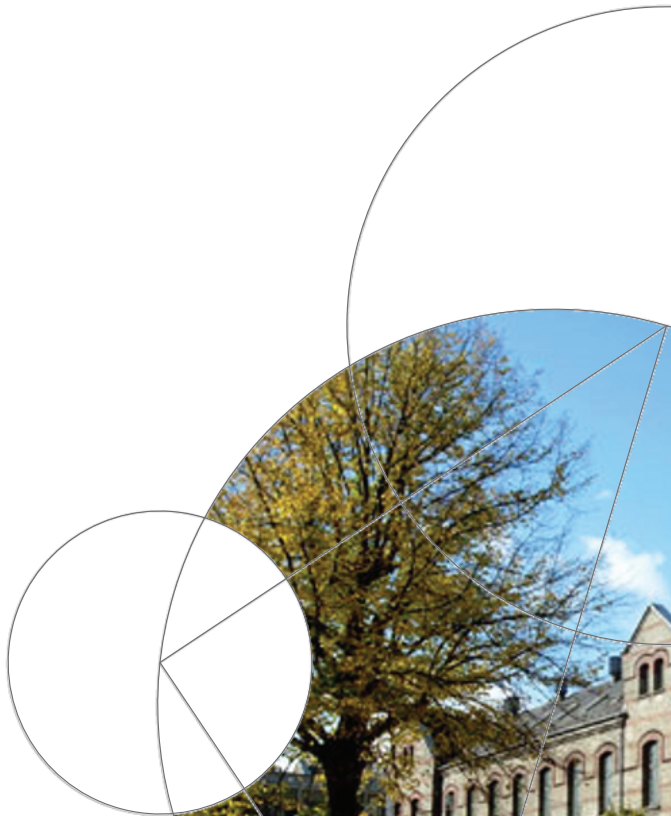
Faculty of Health Sciences

SASA '23 Short Course on Causal Inference

Lecture 3 - Doubly Robust Methods



November 2023

Double Robustness

A doubly-robust estimator is an estimator that is consistent for a given estimand when one or more of the models used in the forming of the estimator is correctly specified for confounding, and possibly censoring.

Estmands and notation

Average treatment effect (ATE), the expected difference in an outcome Y had the full population received and not received the exposure.

$$E\{Y(1)\} - E\{Y(0)\}$$

We will only consider a binary exposure X, and a sufficient set of confounders $\boldsymbol{Z}$.

AIPW

A commonly used Doubly robust estimator for the ATE is the Augmented inverse probability weighted AIPW estimator.

$$\frac{1}{n}\sum_{i=1}^{n}\frac{Y_i X_i}{\hat{p}_i} - \frac{\hat{Y}_{i1}(X_i-\hat{p}_i)}{\hat{p}_i} - \frac{Y_i(1-X_i)}{1-\hat{p}_i} - \frac{\hat{Y}_{i0}(X_i-\hat{p}_i)}{1-\hat{p}_i}.$$

Where $\hat{Y}_{ix}$ is the predicted outcome of Y, setting $X = x$ for subject i and $\hat{p}_i$ is the predicted probability of $X = 1$ for subject i. Any consistent estimation method can be used for the Y and X models.

Two Models equals DR

Recently, there seems to be a general misconception that combining an adjusted outcome model and a propensity score model always gives a doubly robust estimator.

This is not true – it matters how you combine them!

IPTW OLS no interaction

Working outcome model:

$$E(Y|X, \boldsymbol{Z}) = \gamma_0 + \beta X + m(\boldsymbol{Z}; \boldsymbol{\gamma})$$

and a working propensity score model:

$$p(X = 1|\boldsymbol{Z}) = g(\boldsymbol{Z}; \boldsymbol{\alpha})$$

From this you form IPTW:

$$W(X, \boldsymbol{Z}; \hat{\boldsymbol{\alpha}}) = \frac{X}{g(\boldsymbol{Z}; \hat{\boldsymbol{\alpha}})} + \frac{1 - X}{1 - g(\boldsymbol{Z}; \hat{\boldsymbol{\alpha}})},$$

And fit the IPTW score equations:

$$\frac{1}{n} \sum_{i=1}^{n} \left[W(X_i, \boldsymbol{Z}_i; \hat{\boldsymbol{\alpha}}) \left\{ \begin{array}{c} 1 \\ X_i \\ m'(\boldsymbol{Z}_i, \boldsymbol{\gamma}) \end{array} \right\} [Y_i - \gamma_0 - \beta X_i - m(\boldsymbol{Z}_i; \boldsymbol{\gamma})] \right] = 0,$$

where $m'(\boldsymbol{Z}, \boldsymbol{\gamma}) = \frac{dm(\boldsymbol{Z}, \boldsymbol{\gamma})}{d\boldsymbol{\gamma}}$.

IPTW OLS no interaction continued

$\hat{\beta}$ is consistent for the ATE if either the outcome model or the propensity score is correctly specified for confounding.

IPTW OLS General

$$E(Y|X, \boldsymbol{Z}) = \gamma_0 + \beta X + m(\boldsymbol{Z}; \boldsymbol{\gamma}) + \gamma_{xz} X Z_1$$

The double robustness still holds, but now we need to standardize (use G-computation)

$$\hat{E}\{\hat{E}(Y|X = x, \boldsymbol{Z})\} = \frac{1}{n} \sum_{i=1}^{n} \hat{\gamma}_0^{ipw*} + \hat{\beta}^{ipw*} x + m(x, \boldsymbol{Z}_i; \hat{\boldsymbol{\gamma}}^{\boldsymbol{ipw*}}) + \hat{\gamma}_{xz}^{ipw*} x Z_{i1}.$$

Then $\hat{E}\{\hat{E}(Y|X = 1, \boldsymbol{Z})\} - \hat{E}\{\hat{E}(Y|X = 0, \boldsymbol{Z})\}$ is consistent for the ATE if either the propensity score or outcome model is correctly specified for confounding.

Canonical link GLM

Error Distribution	Canonical Link
binomial	$q(\mu) = \log(\mu/(1-\mu))$
Gaussian	$q(\mu) = \mu$
gamma	$q(\mu) = 1/\mu$
inverse Gaussian	$q(\mu) = 1/\mu^2$
Poisson	$q(\mu) = \log(\mu)$

Table: Canonical link functions

IPTW Canonical-link GLM

$$q\{E(Y|X,\boldsymbol{Z})\} = \gamma_0 + \beta X + m(X,\boldsymbol{Z};\boldsymbol{\gamma})$$

$$\frac{1}{n}\sum_{i=1}^{n}\left[W(X_i,\boldsymbol{Z}_i;\hat{\boldsymbol{\alpha}}) \left\{ \begin{array}{c} 1 \\ X_i \\ m'(\boldsymbol{Z}_i,\boldsymbol{\gamma}) \end{array} \right\} [Y_i - q^{-1}\{\gamma_0 + \beta X_i + m(\boldsymbol{Z}_i;\boldsymbol{\gamma})\}] \right] = 0$$

$$\hat{E}\{\hat{E}(Y|X=x,\boldsymbol{Z})\} = \\ \frac{1}{n}\sum_{i=1}^{n} q^{-1}(\hat{\gamma}_0^{ipw**} + \hat{\beta}^{ipw**}x + m(\boldsymbol{Z}_i;\hat{\boldsymbol{\gamma}}^{ipw**})).$$

$$\hat{E}\{\hat{E}(Y|X=1,\boldsymbol{Z})\} - \hat{E}\{\hat{E}(Y|X=0,\boldsymbol{Z})\}$$

Non-canonical link

For a canonical link glm the score equations are always of the form

$$X^T\{y - q^{-1}(X\beta)\} = 0$$

The

$$X^T$$

is the needed piece for the IPTW to cancel out within the weighted estimating equation set-up.
A Non-canonical link glm will not always have this form, thus there is no guarantee that IPT weighting will make a Non-canonical link GLM doubly robust.

Summary for GLM

Working outcome model	Fitting Method	DR
linear regression no interaction	IPTW + OLS	Yes
linear regression	IPTW + OLS + standardization	Yes
canonical link GLM	IPTW + MLE + standardization	Yes
non-canonical link GLM	IPTW + MLE	not generally
non-canonical link GLM	IPTW + MLE + standardization	not generally

All of this is outlined in detail:

Inverse probability of treatment weighting with generalized linear outcome models for doubly robust estimation

`https://arxiv.org/abs/2309.05531`

Summary Recommendations

Do not construct ad hoc estimators!

- Always use canonical link when using IPTW GLM
- Always use canonical link when using GLM in RCT (no IPTW required, but still there is robustness to misspecification)
- We are only considering specification up to the mean model here, so any GLM that has the correct mean model can be used; thus, Poisson for risk ratios
- If you need to combine two models and want DR and are unsure, try AIPW methods (but again, be careful!)

Unfortunately, it also doesn't work in survival

Even after standardization (G-computation) for the survival difference at the time t

- IPTW Cox models, are not doubly robust
- IPTW parametric PH models, are not doubly robust

Away from the null! under the null they are doubly robust

Misspecification, one more time!

Correctly specified for censoring - A correctly specified model that contains a sufficient set of covariates such that, conditional on them, the event time and the censoring time are independent.

A model can be Correctly Specified, but not correctly specified for confounding or censoring, or neither.

a Cox model can be correctly specified, but not for censoring or confounding.

Estimands

Let $T(x)$ be the counterfactual event time under $X = x$. We will assume that the true hazard for the failure time T is given by $\lambda^\star(t|X, \boldsymbol{Z}; \boldsymbol{\omega})$ making the causal conditional log hazard ratio at a given time point t for the true $\boldsymbol{\omega}$

$$\psi(t, \boldsymbol{Z}; \boldsymbol{\omega}) = \log \left\{ \frac{\lambda^\star(t|1, \boldsymbol{Z}; \boldsymbol{\omega})}{\lambda^\star(t|0, \boldsymbol{Z}; \boldsymbol{\omega})} \right\}.$$

This may or may not depend on time, and that $T \perp\!\!\!\perp X|Z$, thus $\psi(t, \boldsymbol{Z}; \boldsymbol{\omega})$ can be considered the causal conditional hazard ratio.

We also consider the causal survival difference at time t
$\zeta(t) = p(T(1) > t) - p(T(0) > t)$

Correctly specified outcome model

In survival PH models, the outcome model needs to be correctly specified for confounding and censoring for the estimated conditional HR to be consistent for $\psi(t, \boldsymbol{Z}; \boldsymbol{\omega})$ the true causal conditional log hazard ratio, away from the null.

However, unlike canonical-link GLM, G-computation doesn't help here; we will still not have a consistent estimator of the difference in survival probabilities, $\zeta(t)$, using IPTW Cox models after standardization away from the null.

DR IPTW Cox under the null

If censoring depends only on the exposure or only of the confounders, and not their interaction, then under null when there is truly no causal effect of the exposure of the survival outcome, then IPT weighting a Cox model will provide a consistent estimate of zero, both in the conditional HR and in the g-computation estimate of the survival difference.

Thus, this can be used as a doubly robust log-rank test. **you need to get the variance estimates correct, use bootstrap!**