Doubly robust estimation

This Lab illustrates notions and concepts presented during the thrid class, with a focus on doubly robustness. Robustness is a statistical property that ensures consistency, even if some of the assumptions are no longer true. In this notebook you will illustrate this property with simulated data. Then, going toward the usage of more flexible tools such as random forest to estimate either the outcome model or the propensity score, we will illustrate the importance of cross-fitting.

All along this lab, n=1000 for sample size is enough to observe the phenomenon we are interested in without taking too computation ressources and time.

Double-robustness with parametric AIPW

1. Implement the parametric AIPW estimator (with logistic regression and regression) and apply it on one simulation from Lunceford. You can also compute a confidence interval.
2. Repeat the process 100 times and compare the results to the ones obtained with the IPW estimator and the g-estimator. Comment.
3. We propose a slight different simulation, where the generative model is close to the one from Lunceford, except that it contains an interaction term or a squared term rather than a fully linear model in the covariate. The function to generate the data is modified in syntheticsimulations. You can use the option mis.prop = TRUE and/or , mis.outcome = TRUE to generate such situations.
4. Simulate the 4 patterns: 1) outcome and propensity score well specified, 2) outcome well specified but propensity score mispecified, 3) outcome misspecified and propensity score well specified, 4) both models are mispecified. Run the estimators, AIPW, IPW and G-estimator on each case. Repeat the process 100 times and represent the results with boxplots. Comment.

AIPW with random forest and cross-fitting

Then, another approach to avoid assumptions is to use flexible model, such as random forest.

The function with cross-fitting is given on the template notebook

Illustrate the results on simulations with non-linear relationship as given in the function simulation wager.