

Summary of ATE Estimators: Pros and Cons

From Whiteboard Session

This document summarizes the pros and cons of various Average Treatment Effect (ATE) estimators as transcribed from the whiteboard screenshots.

| Estimator | Pros | Cons |
|-------------------------------------|--|---|
| Linear Regression Adjustment | <ul style="list-style-type: none">• Simplicity• No tuning required• Interpretable• Provides confidence intervals (CIs)• Very precise if the Conditional Expectation Function (CEF) is truly linear | <ul style="list-style-type: none">• Too simple; likely misspecified• Assumes linear CEF• Problematic if overlap is poor (extrapolation) |
| G-formula (S-learner) | <ul style="list-style-type: none">• Can capture non-linearities | <ul style="list-style-type: none">• Slow convergence rate• Can be too confident (overfitting) if not cross-fit• No straightforward confidence intervals |
| G-formula (T-learner) | <ul style="list-style-type: none">• Often better bias properties than S-learner• Does not “regularize” the treatment | <ul style="list-style-type: none">• Can have worse accuracy for the minority treatment group in imbalanced datasets• No straightforward confidence intervals |

| Estimator | Pros | Cons |
|--------------------|---|--|
| IPW (ML) | <ul style="list-style-type: none"> • Only requires learning the propensity score, not the outcome model • Good accuracy if treatment assignment mechanism is simple and easy to learn from data | <ul style="list-style-type: none"> • High variance • Very sensitive to poor overlap (extreme propensity scores) • No straightforward confidence intervals |
| IPW (Logistic) | <ul style="list-style-type: none"> • Simpler than ML-based IPW • Provides confidence intervals (CIs) (through delta method; which we have not seen in class) | <ul style="list-style-type: none"> • Assumes a simple (logistic) form for the propensity score model |
| Doubly Robust (DR) | <ul style="list-style-type: none"> • Double robustness: Consistent if either the outcome or propensity model is correct • Lower variance than IPW • Asymptotically normal even though ML is used for estimation • Provides confidence intervals (CIs) even though ML is used • Insensitive to errors of ML models | <ul style="list-style-type: none"> • Cannot empirically test the product rate condition for convergence • More randomness due to k-fold cross-validation (can be ameliorated by repeating the process multiple times with different k-folds and taking the median estimate and the median standard error across runs) • Still suffers from substantial (albeit unavoidable without further assumptions) variance if there are covariate regions with low overlap (extreme propensities) |

| Estimator | Pros | Cons |
|-----------------------------|--|--|
| DR (Semi-parametric) | <ul style="list-style-type: none"> • Better computational performance • Better for tuning | <ul style="list-style-type: none"> • Less guardrails against overfitting • If you overdo it with tuning (i.e. let automl run for too long and examine too many models), then you are susceptible to some bias due to overfitting. |
| DR (Stacked) | <ul style="list-style-type: none"> • Reduce variance of tuning • Improve RMSE performance of outcome and propensity model, due to ensembling | <ul style="list-style-type: none"> • Even less guardrails against overfitting • You have to be careful to stack (ensemble) only among a small set of models (e.g. best model of each model type, e.g. (best random forest, best gradient boosted forest, best penalized linear model)) |