Invited Discussion on Bayesian Causal Forests

Arman Oganisian[1] and Jason A. Roy[2]

[1]University of Pennsylvania
[2]Rutgers University

Bayesian Analysis Webinar
September 25, 2020

General Comments on BCF

Many contributions of this work:

- Clear discussion of regularization-induced confounding (RIC) within a nonparametric context.
- Nice illustrations of why targeted selection can be problematic if the propensity score is not included.
- Different BART priors for two functions.

General Comments on BCF

Tempting to stratify and specify priors for $f(x_i, Z_i = 0)$ and $f(x_i, Z_i = 1)$.

- Drawback: lack of direct control over prior for causal effects

Hahn et al. make compelling case for the model:
$f(x_i, z_i) = \mu(x_i) + \tau(x_i) z_i$

- Separating out prognostic score and treatment effect clearly.
- Flexible, yet interpretable shrinkage towards homogeneous effects.

General Comments on BCF

If we start with $f(x_i, z_i) = \mu(x_i) + \tau(x_i)z_i$

- could specify any priors over functions for $\mu()$ and $\tau()$

BCF is more specific. They propose

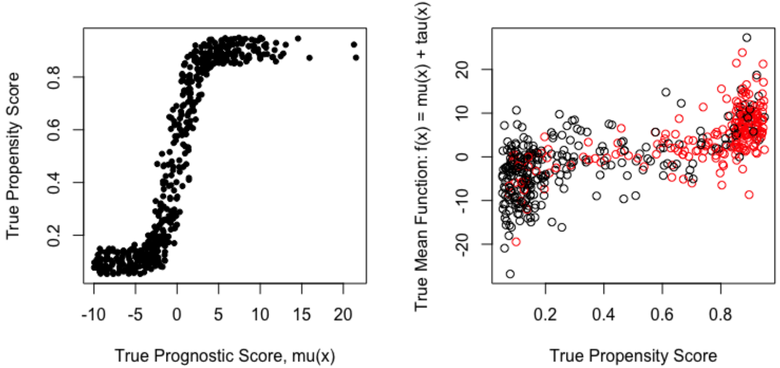
$$f(x_i, z_i) = \mu(x_i, \widehat{\pi}(x_i)) + \tau(x_i)z_i$$

with BART priors on the functions, where $\widehat{\pi}$ is the propensity score.

- Is BART the right choice?
- Why include the propensity score?

Targeted Selection

Targeted selection occurs when the treatment probability depends heavily on the prognostic score (risk if untreated). An example of this is in one of their simulation scenarios:



Targeted Selection

However, ideally selection should be based on the expected benefit of treatment $E(Y(1) - Y(0)|X)$.

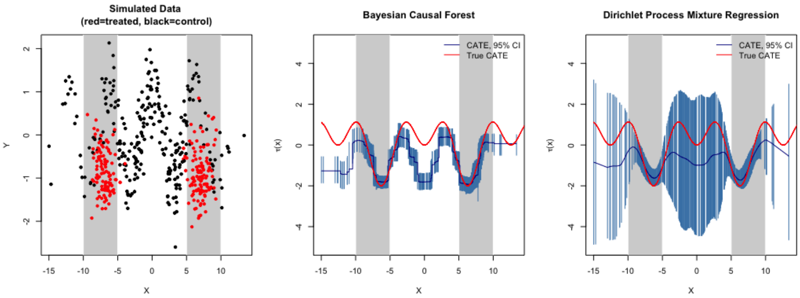
e.g., a clinician should recommend a treatment to patients that they suspect would benefit most from treatment, rather simply basing recommendations on who is the most (or least) frail

How common is targeted selection in practice?

Uncertainty Estimation in Non-overlap Regions

- Overlap must hold: $P(Z \mid X)$ be bounded $\forall X$.
- Model extrapolates in nonoverlap regions where $P(Z \mid X) \approx 1$.
- Trade-off between ignorability and overlap.

Uncertainty Estimation in Non-overlap Regions



BCF inherits some features of BART:

- Non-smooth.
- Homoskedastic.

Some Options

- Ignore it:
  - Underestimates uncertainty (and bias).
- Trimming:
  - Not properly Bayes.
  - For ITE, we "give up" on subjects.
  - For ATE, changes estimand.
- Modified BART ?
  - Smoothed BART [Linero and Yang, 2018].
  - BART with DP prior on errors [George et al., 2018].

References I

George, E., Laud, P., Logan, B., McCulloch, R., and Sparapani, R. (2018).
Fully nonparametric bayesian additive regression trees.

Linero, A. R. and Yang, Y. (2018).
Bayesian regression tree ensembles that adapt to smoothness and sparsity.
Journal of the Royal Statistical Society: Series B (Statistical Methodology), 80(5):1087–1110.