Baseline Simulation

Max Welz
welz@ese.eur.nl

Econometric Institute
Erasmus School of Economics

September 1, 2021

We here describe a baseline simulation design. This simulation design is intended to be as simple as possible. The rationale is that if a certain estimator does not perform well in this simple framework, it will most likely also not perform well in more complicated frameworks. Concretely, this simulation design entails:

- No treatment effect heterogeneity;
- No noise;
- Randomized experiment;
- Perfect information;
- Linear data generating process;
- Identically and independently distributed Gaussian covariates;
- All deceasing individuals die at the exact same time.

Sampling of covariates. Let n be the number of observations and p be the number of covariates. The p-dimensional covariate vector of observation i, X_i, is standard normally distributed. The binary treatment assignment variable W_i assigns the treatment randomly. Mathematically, for all $i = 1, \dots, n$,

$$X_i \sim \mathcal{N}_p(0, I) \quad \text{and} \quad W_i \sim \mathsf{Bernoulli}(0.5).$$

Predictive function. Suppose that $\beta_0 \in \mathbb{R}$ is an intercept term and that $\beta \in \mathbb{R}^p$ is a vector of coefficients associated with the covariates. Let $\theta \in \mathbb{R}$ be the *treatment effect parameter*, which governs the strength of the treatment effect. Since θ is a constant, there is no treatment effect heterogeneity. An effective treatment strategy W_i is characterized by $\theta \neq 0$. We assume that the *predictive function*, $\eta_i(W_i)$, a function of W_i, obeys

$$\eta_i(W_i) := \eta(W_i, X_i) = \beta_0 + X_i^\top \beta + \theta W_i.$$

Hence, the predictive function is assumed to be linear in its parameters (β_0, β, θ).

Mortality risk. We emulate a medical trial. We are interested in the effect of treatment W_i on the (binary) mortality Y_i, where $Y_i = 1$ if individual i dies during the trial, and $Y_i = 0$ if it survives. The probability that individual i with treatment assignment status W_i dies is called the *mortality risk*, which is denoted by $P_i(W_i)$ and defined as

$$P_i(W_i) := P(W_i, X_i) = \mathbb{P}[Y_i = 1 | W_i, X_i] = F_{logistic}\Big(\eta_i(W_i)\Big),$$

where $F_{logistic}$ is the distribution function of the logistic distribution.

Potential Outcomes. The potential outcomes $Y_i(1)$ and $Y_i(0)$ denote the outcome if individual i is treated and if it is *not* treated, respectively. Since the outcome is binary, we assume that

$$Y_i(W_i) \sim \mathsf{Bernoulli}\Big(P_i(W_i)\Big).$$

The *observed outcome* Y_i is defined by

$$Y_i = \begin{cases} Y_i(1) & \text{if } W_i = 1, \\ Y_i(0) & \text{if } W_i = 0. \end{cases}$$

We assume that Y_i is a mortality indicator, that is, $Y_i = 1$ if individual i dies during the time-at-risk (see below for details), and $Y_i = 0$ otherwise.

Time at risk. We assume that the trial takes 10 years and that all deceasing individuals die after 7 years. That is, individual i's *time-at-risk*, T_i, is given by

$$T_i = 7 + 3\mathbb{1}_{\{Y_i = 0\}}.$$

Quantities of interest. We are interested in estimating the following quantities by using the observations $\{(X_i, Y_i, W_i, T_i)\}_{i=1}^n$:

- Treatment effect parameter θ and 95% coverage thereof;
- Average treatment effect: $\mathsf{ATE} = \frac{1}{n}\sum_{i=1}^n \big(Y_i(1) - Y_i(0)\big)$ and 95% coverage thereof;
- Average relative treatment effect: $\mathsf{ARTE} = \frac{\sum_{i=1}^n Y_i(1)}{\sum_{i=1}^n Y_i(0)}$ and 95% coverage thereof;
- Calibration curve, both relative and absolute.

Parameter choices. We fix the parameters as follows.

- $n = 10,000$ and $p = 10$;
- $\theta = -2.5$ (that is, the treatment is effective);

- $\beta_0 = 0$, and, for $j = 1, \ldots, p$,

$$\beta_j = \begin{cases} 1 & \text{if } j \text{ is even} \\ -1 & \text{if } j \text{ is odd.} \end{cases}$$

  Hence, β obeys a Rademacher-like scheme.

These choices lead to the "true" values $\mathsf{ATE} \approx -0.25$ and $\mathsf{ARTE} \approx 0.5$.

Considered models. We evaluate the following models' performance in estimating the quantities of interest.

- one-variable-at-a-time analysis (can only estimate the ARTE);
- risk models and effect models, both with and without survival (in survival models, we effectively fit Cox Proportional Hazard models), as in Kent et al. (2020);
- causal random forest, both with and without survival (Athey et al., 2019 and Cui et al., 2021);
- Double Machine Learning (Chernozhukov et al., 2018) (can only estimate the ATE);
- Generic Machine Learning (Chernozhukov et al., 2020) (give qualitative evaluation of presence of treatment effect heterogeneity).

References

Athey, S., Tibshirani, J., and Wager, S. (2019). Generalized Random Forests. *The Annals of Statistics*, 47(2):1148–1178.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., and Robins, J. (2018). Double/Debiased Machine Learning for Treatment and Structural Parameters. *The Econometrics Journal*, 21(1):C1–C68.

Chernozhukov, V., Demirer, M., Duflo, E., and Fernández-Val, I. (2020). Generic machine learning inference on heterogenous treatment effects in randomized experiments. *arXiv preprint: arXiv1712.04802*.

Cui, Y., Kosorok, M. R., Sverdrup, E., Wager, S., and Zhu, R. (2021). Estimating heterogeneous treatment effects with right-censored data via causal survival forests. *arXiv preprint: arXiv2001.09887*.

Kent, D. M., Paulus, J. K., Van Klaveren, D., D'Agostino, R., Goodman, S., Hayward, R., Ioannidis, J. P., Patrick-Lake, B., Morton, S., Pencina, M., et al. (2020). The predictive approaches to treatment effect heterogeneity (PATH) statement. *Annals of Internal Medicine*, 172(1):35–45.