SPRINT and ACCORD-BP simulators

Address correspondence regarding simulators to Scott Fleming
scottyf@stanford.edu

Draft version March 2022

1 Introduction

In our paper, "Evaluating Treatment Prioritization Rules via Rank-Weighted Average Treatment Effects", we demonstrate how one might use the RATE to evaluate treatment prioritization rules learned from one randomized controlled trial's data (e.g., the Systolic Blood Pressure Intervention Trial (SPRINT)) and applied to a separate randomized trial (e.g., the Action to Control Cardiovascular Risk in Diabetes Blood Pressure trial (ACCORD-BP)). As discussed in the README associated with the paper's code repository, data for the SPRINT and ACCORD-BP clinical trials are available to researchers free of charge via the NHLBI Biologic Specimen and Data Repository Information Coordinating Center (BioLINCC, https://biolincc.nhlbi.nih.gov/home/). However, researchers must submit an online data request form, including a study plan/protocol and Institutional Review Board (IRB) approval, to obtain access. We recognize that this may be prohibitive for some researchers, such as independent investigators not affiliated with an academic research center. In order to make our analysis more accessible to the community, we constructed simulators for both the SPRINT and ACCORD-BP trial data.

2 Simulator Details

2.1 Subject Covariates

In order to simulate data that reflects the first- and second-order moments of the SPRINT and ACCORD-BP randomized controlled trials, we sampled 17 covariates for each subject using an autoregressive procedure parameterized so that the overall covariate distribution approximates the observed data distribution from each trial:

$$
\begin{aligned}
X_{i1} &= g_1(\varepsilon_1; \theta_1) \\
X_{i2} &= g_2(X_{i1}, \varepsilon_2; \theta_2) \\
X_{i3} &= g_3(X_{i1}, X_{i2}, \varepsilon_3; \theta_3) \\
&\vdots \\
X_{ip} &= g_p(X_{i1}, X_{i2}, \dots X_{ip-1}, \varepsilon_p; \theta_p)
\end{aligned}
$$

where ε_i are i.i.d. covariate-specific noise random variables whose exact distribution depends on distribution of the data generating distribution, g_i, and θ_i are covariate-specific autoregressive model parameters.

We note that this procedure is designed not to capture causal relationships between covariates but rather the first- and second-order moments of the observed SPRINT and ACCORD-BP data distributions (i.e., so that the expectation of each covariate and the joint covariance matrix of these covariates in our simulated data approximately match those observed in the actual SPRINT/ACCORD-BP data).

2.2 Censoring Times

In order to simulate censoring times that follow those observed in the SPRINT RCT data, we model the cumulative hazard for a subject being censored as

$$H_C(t) = \left(\frac{t}{1297.74}\right)^{5.05}$$

The censoring times for the ACCORD-BP trial are slightly more difficult to model because, unlike SPRINT, the ACCORD-BP study ran to completion and so censoring rose sharply after three years. The general modeling approach, however, is similar to that of the SPRINT trial (i.e., modeling a cumulative hazard of censoring for each subject). We refer the interested reader to the simulation code for more details.

2.3 Failure Times

We then model the log cumulative hazard of failure for each subject as a function of patient-specific covariates and first-order treatment interactions:

$$\log \lambda(X_i) = \beta_0 + \sum_{j=1}^{p} \beta_j X_{ij} + \sum_{j'=p+1}^{2p} W_i \beta_{j'} X_{ij'}$$

Together with a factor, ρ, estimated from the respective data distributions, we model the overall cumulative hazard of failure at time t as

$$H_T(t) = \left(\frac{t}{\lambda}\right)^{\rho}$$

Finally, we sample the censoring time, C_i, and failure time, T_i, for each subject i using inverse transform sampling. These simulated results are returned to the user for analysis.