simADHDsmart

Description

This function simulates synthetic data to mirror desired characteristics of the Pelham ADHD SMART study. Created specifically for CATIE 2023.

Usage

```
simADHDsmart(
  N = 150,
  baseline.params = list(p.odd = 0.4, m.severity = 0, p.priormed = 0.3, p.race = 0.8),
  Y0.coef = c(2, odd = -0.2, severity = -0.3),
  U.params = c(mu = 0, sd = 1),
  R.coef = c(-0.4, A1 = -0.1, `priormed:A1` = -0.2, U = 0.2),
  adherence.coef = c(-0.1, `priormed:A1` = -0.2, U = 0.1),
  event_time.coef = NULL,
  Y1.coef = c(2.5, A1 = -0.3, U = 0.1),
  Y2.baseline = c(3, odd = -0.3, severity = -0.4, priormed = 0, race = 0.5),
  Y2.tx1 = c(A1 = 0.3, `priormed:A1` = -1.4),
  Y2.n1 = c(R.resid = 0.8, adherence.resid = 1, U = 0.4),
  Y2.tx2 = c(A2 = -0.3, `A1:A2` = 0.1, `adherence:A2` = 1.2),
  sigma = 1
)
```

Arguments

`N`
: number of observations to generate

`baseline.params`
: A list specifying the probability `odd`, mean `severity`, prob. `priormed`, prob. `race`

`Y0.coef`
: A named vector specifying the linear model coefficients for baseline school performance. Can be a function of any past variables. See *Named Vectors*

`U.params`
: A vector specifying the normal distribution parameters `mu` and `sd`

`R.coef`
: A named vector specifying the linear **probit model** coefficients for the probability of being a responder to first-stage treatment. Can be a function of any past variables.

`adherence.coef`
: A named vector specifying the linear **probit model** coefficients for the probability of being adherent to first-stage treatment. Can be a function of any past variables.

`event_time.coef` not implemented

`Y1.coef`
: A named vector of linear model coefficients specifying the first-stage treatment causal effect on Y_1 school performance

`Y2.baseline`
: A named vector of linear model coefficients specifying the baseline covariate associations on end-of-study Y_2 school performance

`Y2.tx1`
: A named vector of linear model coefficients specifying the first-stage treatment causal effect on end-of-study Y_2 school performance. Can be a function of any baseline moderators.

`Y2.n1`
: A named vector of linear model coefficients specifying the nuisance associations on end-of-study Y_2 school performance. Can be a function of any prior moderators. NOTE: must specify R.resid and adherence.resid which are the direct associations in the SNMM.

`Y2.tx2`
: A named vector of linear model coefficients specifying the second-stage treatment causal effect, among non-responders, on end-of-study Y_2 school performance. Can be a function of any baseline moderators. NOTE: all moderators are grand mean centered

`sigma`
: gaussian noise added to $Y_{0,1,2}$

Value

A list with components

data

: data.frame of observed variables

DTRmean

: marginal embedded DTR means

Named Vectors

The named vectors of coefficients use formula notation to specify interactions i.e. "`priormed:A1`", which must be quoted. The order does not matter but spelling does. Note: if the first element is unamed it is treated as the intercept term, else if all elements are named the intercept is ommited.

Operating Characteristics

All baseline and moderator variable are grand mean centered. The nuisance terms `R.resid` and `adherence.resid` are residualized using the true probabilities.

Default Structural Nested Mean Model:

School performance after first-stage

$$Y_1(a_1, a_2) = 2.5 - 0.3a_1 + 0.4U + \epsilon$$

School performance after second-stage

$$\begin{gathered} Y_2(a_1, a_2) = 3 - 0.2odd - 0.3severity + 0 * priormed + .5race + \\ (0.3 - 1.4priormed)a_1 + \\ 0.8R.resid(a_1, u) + 1adherence.resid(a_1, u) \\ +(1 - R)(-0.3 + 0.1a_1 + 1.2adherence)a_2 + 0.4U + \epsilon \end{gathered}$$

Baseline Covariates:

- `odd`: binary, centered
- `severity`: standard normal
- `priormed`: binary, centered
- `race`: binary, centered
- `U`: an unknown, common cause of `R`, `adherence`, `Y1`, and `Y2`. Induces collider bias if naively conditioning on the time-varying covariates.

First-Stage:

- Marginally, `BMOD(1)` is initially worse compared to `MED(-1)` after the first-stage Y_1, but better in the long run Y_2
- IF on `priormed`, starting with `MED` is better

Nuisance Associations

- `R`: positively associated with outcome
- `adherence`: positively associated with outcome
- `event-time`: or time to non-response has no effect

Second-Stage (among non-responders only):

- Marginally, `AUG(-1)` is better compared to `INT(1)`

- Positive interaction for `AUG` if given `MED`, and `INT` if given `BMOD`
- IF adherent to first-stage, much better to `INT`; if non-adherent to first-stage, much better to `AUG`

Internal functions

- linearMult: takes a named vector of coefficients and a data.frame and creates the necessary design matrix to multiply and return the function values.
- sampleProbitMean: generates observations from the posterior in order to empirically evaluate probability the integral.
- getMarginalMeans: takes the specified coefficients and returns the marginal means of the four embedded adaptive interventions, averaging over response and adherence.

Author(s)

Mason Ferlic, December 2022, University of Michigan, D3C
