

Package ‘borrowr’

December 8, 2020

Type Package

Title Estimate Causal Effects with Borrowing Between Data Sources

Version 0.2.0

Author Jeffrey A. Boatman [aut, cre],
David M. Vock [aut],
Joseph S. Koopmeiners [aut]

Maintainer Jeffrey A. Boatman <jeffrey.boatman@gmail.com>

Description Estimate population average treatment effects from a primary data source with borrowing from supplemental sources. Causal estimation is done with either a Bayesian linear model or with Bayesian additive regression trees (BART) to adjust for confounding. Borrowing is done with multisource exchangeability models (MEMs). For information on BART, see Chipman, George, & McCulloch (2010) <doi:10.1214/09-AOAS285>. For information on MEMs, see Kaizer, Koopmeiners, & Hobbs (2018) <doi:10.1093/biostatistics/kxx031>.

Depends R (>= 3.5.0)

License GPL (>= 3)

Encoding UTF-8

LazyData true

RoxygenNote 7.1.1

Imports mvtnorm(>= 1.0.8), BART(>= 2.1), Rcpp (>= 1.0.0)

LinkingTo Rcpp

Suggests knitr, rmarkdown, ggplot2

VignetteBuilder knitr

NeedsCompilation yes

Repository CRAN

Date/Publication 2020-12-08 12:50:03 UTC

R topics documented:

adapt	2
credint	3
pate	3

adapt

Data set used in the package vignette

Description

A simulated data set used in the package vignette. Data generating mechanism is adapted from Hill (2011). Includes 3 data sources to illustrate borrowing information in estimate the population average treatment effect.

Usage

adapt

Format

A data frame with 180 rows and 5 variables

y the outcome

x a variable associated with y and treatment

source character variable with 3 levels: "Primary", "Supp1", "Supp2"

treatment treatment variable. 1 = treated, 0 = untreated

compliant compliance indicator. 1 = compliant to treatment, 0 = noncompliant to treatment

References

Hill, Jennifer L. (2011) Bayesian Nonparametric Modeling for Causal Inference. *Journal of Computational and Graphical Statistics*, 20:1, 217-240.

Examples

```
data(adapt)
head(adapt)
```

credint	<i>Posterior Credible Interval for Population Average Treatment Effect (PATE)</i>
---------	---

Description

Computes an equal-tailed credible interval for an object of class 'pate'

Usage

```
credint(object, level = 0.95)
```

Arguments

object	An object of class pate fit by the pate function.
level	the credible level required

Examples

```
data(adapt)

est <- pate(y ~ treatment*x + treatment*I(x ^ 2), data = adapt,
  estimator = "bayesian_lm", src_var = "source", primary_source = "Primary",
  trt_var = "treatment")

credint(est)
```

pate	<i>Population Average Treatment Effect (PATE)</i>
------	---

Description

Estimates the population average treatment effect from a primary data source with potential borrowing from supplemental sources. Adjust for confounding by fitting a conditional mean model with the treatment variable and confounding variables.

Usage

```
pate(
  formula,
  estimator = c("BART", "bayesian_lm"),
  data,
  src_var,
  primary_source,
  exch_prob,
```

```

    trt_var,
    compliance_var,
    ndpost = 1000,
    model_prior = c("none", "power", "powerlog"),
    ...
  )

```

Arguments

formula	An object of class formula. The left hand side must be the outcome, and the right hand side must include the treatment variable. To adjust for confounding, the right hand side must also include the confounders. For the Bayesian linear model, the user is responsible for specifying the function form.
estimator	"bayesian_lm" or "BART". If "bayesian_lm", a Bayesian linear model with a normal inverse-gamma prior. If "BART", Bayesian Additive Regression Trees
data	the data frame with the outcome variable and all variables in the formula
src_var	a character variable which variable indicates the source variable. Must match a column name in the data.
primary_source	character variable indicating the primary source. Must match one of the values of src_var.
exch_prob	numeric vector giving prior probability that each source is exchangeable with the primary source. Number of elements must be equal to the number of sources minus 1. Each element must be between 0 and 1. Order of probabilities should match the order returned by calling <code>levels(factor(...))</code> on the source vector. See the vignette for an example of usage.
trt_var	which variable indicates the treatment. Must match a column name in the data. Must be coded as numeric values 0 and 1, 0 for untreated, 1 for treated.
compliance_var	Optional argument if adjustment for confounding due to noncompliance is needed. compliance_var indicates the compliance indicator. Must match a column name in the data. Must be coded as numeric values 0 and 1, 0 for noncompliant, 1 for compliant. If this argument is specified, the formula must also include the compliance variable.
ndpost	number of draws from the posterior
model_prior	specifies the prior probability of exchangeability of data sources. Details for priors are given in Boatman et al. (2020).
...	additional arguments passed to BART

Details

To adjust for confounding, the PATE is estimated using a model for the conditional mean given treatment and confounders. Currently, two models are available, a Bayesian linear model with an inverse-gamma prior, and Bayesian Additive Regression Trees (BART; Chipman & McCulloch, 2010). The user must specify a formula for the conditional mean. This requires more thought for the Bayesian linear model as the analyst must carefully consider the functional form of the regression relationship. For BART, the right hand side of the formula need only include the confounders and

the treatment variable without specification of the functional form. If there is no confounding, the right hand side of the formula needs to include the treatment variable only.

If `formula = "bayesian_lm"`, then the function fits the Bayesian linear model

$$Y = X\beta + \epsilon, \epsilon \sim N(0, \sigma^2).$$

The prior on the regression coefficients is normal with mean vector 0 and variance matrix with diagonal elements equal to 100 and off-diagonal elements equal to 0. The prior on σ^2 is a Inverse Gamma(0.1, 0.1)

If `formula = "BART"`, the function fits the Bayesian Additive Regression Trees model, but with a modified prior on the terminal nodes. The prior on each terminal node is

$$N(0, \gamma\sigma^2).$$

The package uses the default value

$$\gamma = 1/(16 * m * \hat{\sigma}^2)$$

where m is the number of trees and $\hat{\sigma}^2$ is the variance of Y .

Borrowing between data sources is done with Multisource Exchangeability Models (MEMs; Kaizer et al., 2018). MEMs borrow by assuming that each supplementary data source is either "exchangeable", or not, with the primary data source. Two data sources are considered exchangeable if their model parameters are equal. Each data source can be exchangeable with the primary data, or not, so if there are r data sources, there are 2^r possible configurations regarding the exchangeability assumptions. Each of these configurations corresponds to a single MEM. The parameters for each MEM are estimated, and we compute a posterior probability for each. The posterior density of the PATE is a weighted posterior across all possible MEMs.

Value

A list with components:

<code>call</code>	The function call
<code>estimator</code>	The estimator used to adjust for confounding, "bayesian_lm" for Bayesian linear model, or "BART" for Bayesian additive regression trees
<code>EY0</code>	Posterior draws of the expected potential outcome if all observations were treated. One column for each MEM
<code>EY1</code>	Posterior draws of the expected potential outcome if all observations were untreated. One column for each MEM
<code>log_marg_like</code>	Log marginal likelihood for each MEM
<code>mem_pate_post</code>	Array containing, for each MEM, posterior draws for the population average treatment effect
<code>MEMs</code>	Matrix showing showing which data sources are exchangeable under each MEM. 1 = exchangeable.
<code>pate_post</code>	Posterior draws for the population average treatment effect. Weighted average of <code>mem_pate_post</code> , where <code>post_probs</code> are the weights

post_probs	Posterior probability that each MEM (shown in the list element MEMs) is the true model.
exch_prob	Prior probability that each source is exchangeable with the primary source.
beta_post_mean	If estimator = "bayesian_lm", a matrix with the posterior means of the coefficients for the primary source from each MEM. If estimator = "BART", NA
.	.
beta_post_var	If estimator = "bayesian_lm", a matrix with the posterior variance of the coefficients for the primary source from each MEM. If estimator = "BART", NA
.	.
beta_post_var	If estimator = "bayesian_lm", a matrix with the posterior variance of the coefficients for the primary source from each MEM. If estimator = "BART", NA
.	.
non_compliance	Logical, indicating whether the model adjusted for confounding due to noncompliance

References

- Chipman, H. & McCulloch, R. (2010) BART: Bayesian additive regression trees. *Annals of Applied Statistics*, 4(1): 266-298.
- Kaizer, Alexander M., Koopmeiners, Joseph S., Hobbs, Brian P. (2018) Bayesian hierarchical modeling based on multisource exchangeability. *Biostatistics*, 19(2): 169-184.
- Boatman, Jeffrey A., Vock, David. M, and Koopmeiners, Joseph S. (2020) Borrowing from Supplemental Sources to Estimate Causal Effects from a Primary Data Source. [arXiv:2003.09680](https://arxiv.org/abs/2003.09680)

Examples

```
data(adapt)

est <- pate(y ~ treatment*x + treatment*I(x ^ 2), data = adapt,
  estimator = "bayesian_lm", src_var = "source", primary_source = "Primary",
  trt_var = "treatment")

summary(est)
```

Index

* **datasets**
 adapt, 2

adapt, 2

credint, 3

pate, 3