Package ‘npcausal’

May 17, 2017

Type Package

Title Nonparametric causal inference methods

Version 0.1.0

Author Edward H. Kennedy

Maintainer Edward H. Kennedy <edward@stat.cmu.edu>

Description This package provides a variety of tools for nonparametric estimation of causal effects across a wide range of settings. The methods are based on the theory of influence functions, and can incorporate flexible machine learning and high-dimensional regression tools, while still yielding inference in the form of confidence intervals and hypothesis tests. Many of the methods are doubly robust.

License GPL

Encoding UTF-8

LazyData true

RoxygenNote 6.0.1

R topics documented:

`ate` *Estimating average effect of discrete treatment*

Description

`ate` is used to estimate the mean outcome in a population had all subjects received given levels of a discrete (unconfounded) treatment.

Usage

```
ate(x, a, y, nsplits=2,
sl.lib=c("SL.earth","SL.gam","SL.glm","SL.glmnet", "SL.glm.interaction","SL.mean","SL.ranger"))
```

Arguments

x covariate matrix.

a discrete treatment.

y outcome of interest.

nsplits integer number of sample splits for nuisance estimation. If nsplits=1, sample splitting is not used, and nuisance functions are estimated on full sample (in which case validity of SEs/CIs requires empirical process conditions). Otherwise must have nsplits>1.

sl.lib algorithm library if using SuperLearner. Default library includes "earth", "gam", "glm", "glmnet", "glm.interaction", "mean", and "ranger".

Value

A list containing the following components:

res estimates/SEs/CIs/p-values for population means and relevant contrasts.

ifvals matrix of estimated influence function values.

References

Tsiatis AA. *Semiparametric Theory and Missing Data* (Springer, 2006).

Examples

```
n <- 1000; x <- matrix(rnorm(n*4),nrow=n)
a <- sample(4,n,replace=TRUE); y <- rnorm(n)

ate.res <- ate(x,a,y)
```

att *Estimating average effect of treatment on the treated*

Description

att is used to estimate the difference in mean outcome among treated subjects had a binary (unconfounded) treatment been withheld.

Usage

```
att(x, a, y, nsplits=2,
sl.lib=c("SL.earth","SL.gam","SL.glm","SL.glmnet", "SL.glm.interaction","SL.mean","SL.ranger"))
```

Arguments

`x` covariate matrix.

`a` binary treatment.

`y` outcome of interest.

`nsplits` integer number of sample splits for nuisance estimation. If nsplits=1, sample splitting is not used, and nuisance functions are estimated on full sample (in which case validity of SEs/CIs requires empirical process conditions). Otherwise must have nsplits>1.

`sl.lib` algorithm library if using SuperLearner. Default library includes "earth", "gam", "glm", "glmnet", "glm.interaction", "mean", and "ranger".

Value

A list containing the following components:

`res` estimates/SEs/CIs/p-values for treated means and contrast.

`ifvals` vector of estimated influence function values.

References

Hahn J (1998). On the role of the propensity score in efficient semiparametric estimation of average treatment effects. *Econometrica.*

Kennedy EH, Sjolander A, Small DS (2015). Semiparametric causal inference in matched cohort studies. *Biometrika.*

Examples

```
n <- 1000; x <- matrix(rnorm(n*4),nrow=n)
a <- rbinom(n,1,.3); y <- rnorm(n)

att.res <- att(x,a,y)
```

`ctseff` *Estimating average effect curve for continuous treatment*

Description

`ctseff` is used to estimate the mean outcome in a population had all subjects received given levels of a continuous (unconfounded) treatment.

Usage

```
ctseff(x, a, y, nsplits=2,
sl.lib=c("SL.earth","SL.gam","SL.glm","SL.glmnet", "SL.glm.interaction","SL.mean","SL.ranger"))
```

Arguments

`x` covariate matrix.

`a` continuous treatment.

`y` outcome of interest.

`nsplits` integer number of sample splits for nuisance estimation. If nsplits=1, sample splitting is not used, and nuisance functions are estimated on full sample (in which case validity of SEs/CIs requires empirical process conditions). Otherwise must have nsplits>1.

`sl.lib` algorithm library if using SuperLearner. Default library includes "earth", "gam", "glm", "glmnet", "glm.interaction", "mean", and "ranger".

Value

A list containing the following components:

`res` estimates/SEs/CIs/p-values for population means and relevant contrasts.

`ifvals` matrix of estimated influence function values.

References

Kennedy EH, Ma Z, McHugh MD, Small DS (2017). Nonparametric methods for doubly robust estimation of continuous treatment effects. *Journal of the Royal Statistical Society, Series B*. arxiv:1507.00747

Examples

```
n <- 1000; x <- matrix(rnorm(n*4),nrow=n)
a <- runif(n); y <- rnorm(n)

ctseff.res <- ctseff(x,a,y)
```

ipsi *Estimating effects of incremental propensity score interventions*

Description

`ipsi` is used to estimate effects of incremental propensity score interventions.

Usage

```
ipsi(dat, x.trt, x.out, delta.seq, nsplits)
```

Arguments

`dat` dataframe (in long not wide form if longitudinal) with columns ‘time’, ‘id’, outcome ‘y’, treatment ‘a’.

`x.trt` covariate matrix for treatment regression.

`x.out` covariate matrix for outcome regression.

`delta.seq` sequence of delta values.

`nsplits` number of sample splits.

Value

The sum of x and y.

References

Kennedy EH. Nonparametric causal effects based on incremental propensity score interventions. arxiv:1704.00211

Index