

Package ‘npcausal’

June 13, 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	1
att	3
ctseff	4
ipsi	5
plot.ctseff	6
SL.ranger	7
Index	8

ate	<i>Estimating average effect of discrete treatment</i>
-----	--

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

Arguments

y	outcome of interest.
a	discrete treatment.
x	covariate matrix.
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 for SuperLearner. Default library includes "earth", "gam", "glm", "glmnet", "glm.interaction", "mean", "ranger", "rpart".

Value

A list containing the following components:

res	estimates/SEs/CIs/p-values for population means and relevant contrasts.
nuis	subject-specific estimates of nuisance functions (i.e., propensity score and outcome regression)
ifvals	matrix of estimated influence function values.

References

- Robins JM, Rotnitzky A (1995). Semiparametric efficiency in multivariate regression models with missing data. *Journal of the American Statistical Association*.
- Hahn J (1998). On the role of the propensity score in efficient semiparametric estimation of average treatment effects. *Econometrica*.
- van der Laan MJ, Robins JM (2003). *Unified Methods for Censored Longitudinal Data and Causality* (Springer).
- Tsiatis AA (2006). *Semiparametric Theory and Missing Data* (Springer).
- Robins JM, Li L, Tchetgen Tchetgen ET, van der Vaart A (2008). Higher order influence functions and minimax estimation of nonlinear functionals. *Probability and Statistics: Essays in Honor of David A. Freedman*.
- Zheng W, van der Laan (2010). Asymptotic theory for cross-validated targeted maximum likelihood estimation *UC Berkeley Division of Biostatistics Working Paper Series*.
- Chernozhukov V, Chetverikov V, Demirer M, et al (2016). Double machine learning for treatment and causal parameters.

Examples

```
n <- 1000; x <- matrix(rnorm(n*5),nrow=n)
a <- sample(3,n,replace=TRUE); y <- rnorm(n)

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

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

Arguments

y	outcome of interest.
a	binary treatment.
x	covariate matrix.
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.
nuis	subject-specific estimates of nuisance functions (i.e., propensity score and outcome regression)
ifvals	vector of estimated influence function values.

References

(Also see references for function ate)

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

Examples

```
n <- 1000; x <- matrix(rnorm(n*5),nrow=n)
a <- rbinom(n,1,.3); y <- rnorm(n)

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

ctseff

Estimating average effect curve for continuous treatment

Description

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

Usage

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

Arguments

y	outcome of interest.
a	continuous treatment.
x	covariate matrix.
bw.seq	sequence of bandwidth values.
sl.lib	algorithm library for SuperLearner. Default library includes "earth", "gam", "glm", "glmnet", "glm.interaction", "mean", and "ranger".

Value

A list containing the following components:

res	estimates/SEs/CIs for population means.
bw.risk	estimated risk at sequence of bandwidth 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](https://arxiv.org/abs/1507.00747)

Examples

```
n <- 500; x <- matrix(rnorm(n*5),nrow=n)
a <- runif(n); y <- a + rnorm(n,sd=.5)

ce.res <- ctseff(y,a,x, bw.seq=seq(.2,2,length.out=100))
plot.ctseff(ce.res)

# check that bandwidth choice is minimizer
plot(ce.res$bw.risk$bw,ce.res$bw.risk$risk)
```

ipsi

*Estimating effects of incremental propensity score interventions***Description**

ipsi is used to estimate effects of incremental propensity score interventions, i.e., estimates of mean outcomes if the odds of receiving treatment were multiplied by a factor delta.

Usage

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

Arguments

y	outcome of interest measured at end of study.
a	binary treatment.
x.trt	covariate matrix for treatment regression.
x.out	covariate matrix for outcome regression.
time	measurement time.
id	subject identifier.
delta.seq	sequence of delta increment values.
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.

Value

A list containing the following components:

res	estimates/SEs and uniform CIs for population means.
res.ptwise	estimates/SEs and pointwise CIs for population means.
calpha	multiplier bootstrap critical value.

Details

Treatment and covariates are expected to be time-varying and measured throughout the course of the study. Therefore if n is the number of subjects and T the number of timepoints, then `a`, `time`, and `id` should all be vectors of length $n \times T$, and `x.trt` and `x.out` should be matrices with $n \times T$ rows. However `y` should be a vector of length n since it is only measured at the end of the study. The subject ordering should be consistent across function inputs, based on the ordering specified by `id`. See example below for an illustration.

References

Kennedy EH. Nonparametric causal effects based on incremental propensity score interventions. [arxiv:1704.00211](https://arxiv.org/abs/1704.00211)

Examples

```
n <- 500; T <- 4

time <- rep(1:T,n); id <- rep(1:n,rep(T,n))
x.trt <- matrix(rnorm(n*T*5),nrow=n*T)
x.out <- matrix(rnorm(n*T*5),nrow=n*T)
a <- rbinom(n*T,1,.5); y <- rnorm(n)

d.seq <- seq(0.1,5,length.out=10)

ipsi.res <- ipsi(y,a, x.trt,x.out, time,id, d.seq)
```

plot.ctseff

Plot estimated average effect curve for continuous treatment

Description

plot.ctseff is used to plot results from ctseff fit.

Usage

```
plot.ctseff(ctseff.res)
```

Arguments

ctseff.res output from ctseff fit.

Value

A plot of estimated effect curve with pointwise confidence intervals.

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](https://arxiv.org/abs/1507.00747)

Examples

```
n <- 500; x <- matrix(rnorm(n*5),nrow=n)
a <- runif(n); y <- a + rnorm(n,sd=.5)

ce.res <- ctseff(y,a,x, bw.seq=seq(.2,2,length.out=100))
plot.ctseff(ce.res)
```

SL.ranger*Add Ranger wrapper for SuperLearner*

Description

SL.ranger is a wrapper for SuperLearner that adds the fast random forests method ranger.

Usage

```
SL.ranger(Y, X, newX, family, ...)
```

Arguments

Y	outcome vector.
X	covariate dataframe for training.
newX	covariate dataframe for predictions.
family	link function (currently only supports "gaussian" identity link).

Value

Predictions and fits from ranger.

References

Wright MN, Ziegler A (2016). ranger: A fast implementation of random forests for high dimensional data in C++ and R. *Journal of Statistical Software*.

Index

ate, [1](#)

att, [3](#)

ctseff, [4](#)

ipsi, [5](#)

plot.ctseff, [6](#)

SL.ranger, [7](#)