Package 'ATbounds'

January 20, 2025

Type Package

Title Bounding Treatment Effects by Limited Information Pooling

Version 0.1.0

Description Estimation and inference methods for bounding average treatment effects (on the treated) that are valid under an unconfoundedness assumption.
The bounds are designed to be robust in challenging situations, for example, when the conditioning variables take on a large number of different values in the observed sample, or when the overlap condition is violated.
This robustness is achieved by only using limited ``pooling" of information across observations.
For more details, see the paper by Lee and Weidner (2021), ``Bounding Treatment Effects by Pooling Limited Information across Observations," <arXiv:2111.05243>.

License GPL-3

Encoding UTF-8

LazyData true

RoxygenNote 7.1.2

Imports stats, mgcv

Suggests knitr, rmarkdown, testthat, ggplot2

VignetteBuilder knitr

Depends R (>= 2.10)

URL https://github.com/ATbounds/ATbounds-r/

BugReports https://github.com/ATbounds/ATbounds-r/issues

NeedsCompilation no

Author Sokbae Lee [aut, cre],
Martin Weidner [aut]

Maintainer Sokbae Lee <sl3841@columbia.edu>

Repository CRAN

Date/Publication 2021-11-24 19:00:02 UTC

Contents

atebounds 2
attbounds 3
EFM 5
RHC 6
simulation_dgp 9
summary.ATbounds 10

Index **12**

`atebounds` *Bounding the average treatment effect (ATE)*

Description

Bounds the average treatment effect (ATE) under the unconfoundedness assumption without the overlap condition.

Usage

```
atebounds(
  Y,
  D,
  X,
  rps,
  Q = 3L,
  studentize = TRUE,
  alpha = 0.05,
  x_discrete = FALSE,
  n_hc = NULL
)
```

Arguments

`Y`	n-dimensional vector of binary outcomes
`D`	n-dimensional vector of binary treatments
`X`	n by p matrix of covariates
`rps`	n-dimensional vector of the reference propensity score
`Q`	bandwidth parameter that determines the maximum number of observations for pooling information (default: Q = 3)
`studentize`	TRUE if the columns of X are studentized and FALSE if not (default: TRUE)
`alpha`	(1-alpha) nominal coverage probability for the confidence interval of ATE (default: 0.05)
`x_discrete`	TRUE if the distribution of X is discrete and FALSE otherwise (default: FALSE)
`n_hc`	number of hierarchical clusters to discretize non-discrete covariates; relevant only if x_discrete is FALSE. The default choice is n_hc = ceiling(length(Y)/10), so that there are 10 observations in each cluster on average.

Value

An S3 object of type "ATbounds". The object has the following elements.

call	a call in which all of the specified arguments are specified by their full names
type	ATE
cov_prob	Confidence level: 1-alpha
y1_lb	estimate of the lower bound on the average of Y(1), i.e. E[Y(1)]
y1_ub	estimate of the upper bound on the average of Y(1), i.e. E[Y(1)]
y0_lb	estimate of the lower bound on the average of Y(0), i.e. E[Y(0)]
y0_ub	estimate of the upper bound on the average of Y(0), i.e. E[Y(0)]
est_lb	estimate of the lower bound on ATE, i.e. E[Y(1) - Y(0)]
est_ub	estimate of the upper bound on ATE, i.e. E[Y(1) - Y(0)]
est_rps	the point estimate of ATE using the reference propensity score
se_lb	standard error for the estimate of the lower bound on ATE
se_ub	standard error for the estimate of the upper bound on ATE
ci_lb	the lower end point of the confidence interval for ATE
ci_ub	the upper end point of the confidence interval for ATE

References

Sokbae Lee and Martin Weidner. Bounding Treatment Effects by Pooling Limited Information across Observations.

Examples

```
Y <- RHC[,"survival"]
D <- RHC[,"RHC"]
X <- RHC[,c("age","edu")]
rps <- rep(mean(D),length(D))
results_ate <- atebounds(Y, D, X, rps, Q = 3)
```

`attbounds` *Bounding the average treatment effect on the treated (ATT)*

Description

Bounds the average treatment effect on the treated (ATT) under the unconfoundedness assumption without the overlap condition.

Usage

```
attbounds(
  Y,
  D,
  X,
  rps,
  Q = 3L,
  studentize = TRUE,
  alpha = 0.05,
  x_discrete = FALSE,
  n_hc = NULL
)
```

Arguments

Y	n-dimensional vector of binary outcomes
D	n-dimensional vector of binary treatments
X	n by p matrix of covariates
rps	n-dimensional vector of the reference propensity score
Q	bandwidth parameter that determines the maximum number of observations for pooling information (default: Q = 3)
studentize	TRUE if X is studentized elementwise and FALSE if not (default: TRUE)
alpha	(1-alpha) nominal coverage probability for the confidence interval of ATE (default: 0.05)
x_discrete	TRUE if the distribution of X is discrete and FALSE otherwise (default: FALSE)
n_hc	number of hierarchical clusters to discretize non-discrete covariates; relevant only if x_discrete is FALSE. The default choice is n_hc = ceiling(length(Y)/10), so that there are 10 observations in each cluster on average.

Value

An S3 object of type "ATbounds". The object has the following elements.

call	a call in which all of the specified arguments are specified by their full names
type	ATT
cov_prob	Confidence level: 1-alpha
est_lb	estimate of the lower bound on ATT, i.e. E[Y(1) - Y(0) | D = 1]
est_ub	estimate of the upper bound on ATT, i.e. E[Y(1) - Y(0) | D = 1]
est_rps	the point estimate of ATT using the reference propensity score
se_lb	standard error for the estimate of the lower bound on ATT
se_ub	standard error for the estimate of the upper bound on ATT
ci_lb	the lower end point of the confidence interval for ATT
ci_ub	the upper end point of the confidence interval for ATT

References

Sokbae Lee and Martin Weidner. Bounding Treatment Effects by Pooling Limited Information across Observations.

Examples

```
Y <- RHC[,"survival"]
D <- RHC[,"RHC"]
X <- RHC[,c("age","edu")]
rps <- rep(mean(D),length(D))
results_att <- attbounds(Y, D, X, rps, Q = 3)
```

EFM *EFM*

Description

The electronic fetal monitoring (EFM) and cesarean section (CS) dataset from Neutra, Greenland, and Friedman (1980) consists of observations on 14,484 women who delivered at Beth Israel Hospital, Boston from January 1970 to December 1975. The purpose of the study is to evaluate the impact of EFM on cesarean section (CS) rates. It is found by Neutra, Greenland, and Friedman (1980) that relevant confounding factors are: nulliparity (nullipar), arrest of labor progression (arrest), malpresentation (breech), and year of study (year). The dataset provided in the R package is from the supplementary materials of Richardson, Robins, and Wang (2017), who used this dataset to illustrate their proposed methods for modeling and estimating relative risk and risk difference.

Usage

```
EFM
```

Format

A data frame with 14484 rows and 6 variables:

cesarean Outcome: 1 if delivery was via cesarean section; 0 otherwise

monitor Treatment: 1 if electronic fetal monitoring (EFM) was used; 0 otherwise

arrest Covariate: 1 = arrest of labor progression; 0 otherwise

breech Covariate: 1 = malpresentation (breech); 0 otherwise

nullipar Covariate: 1 = nulliparity; 0 otherwise

year Year of study: 0,...,5 (actual values are 1970,...,1975)

Source

The dataset from Neutra, Greenland, and Friedman (1980) is available as part of supplementary materials of Richardson, Robins, and Wang (2017) on Journal of the American Statistical Association website at doi: 10.1080/01621459.2016.1192546.

References

Neutra, R.R., Greenland, S. and Friedman, E.A., 1980. Effect of fetal monitoring on cesarean section rates. Obstetrics and gynecology, 55(2), pp.175-180.

Richardson, T.S., Robins, J.M. and Wang, L., 2017. On modeling and estimation for the relative risk and risk difference. Journal of the American Statistical Association, 112(519), pp.1121-1130.

RHC *RHC*

Description

The right heart catheterization (RHC) dataset is publicly available on the Vanderbilt Biostatistics website. RHC is a diagnostic procedure for directly measuring cardiac function in critically ill patients. The dependent variable is 1 if a patient survived after 30 days of admission, 0 if a patient died within 30 days. The treatment variable is 1 if RHC was applied within 24 hours of admission, and 0 otherwise. The sample size was n = 5735, and 2184 patients were treated with RHC. Connors et al. (1996) used a propensity score matching approach to study the efficacy of RHC, using data from the observational study called SUPPORT (Murphy and Cluff, 1990). Many authors used this dataset subsequently. The 72 covariates are constructed, following Hirano and Imbens (2001).

Usage

```
RHC
```

Format

A data frame with 5735 rows and 74 variables:

survival Outcome: 1 if a patient survived after 30 days of admission, and 0 if a patient died within 30 days

RHC Treatment: 1 if RHC was applied within 24 hours of admission, and 0 otherwise.

age Age in years

edu Years of education

cardiohx Cardiovascular symptoms

chfhx Congestive Heart Failure

dementhx Dementia, stroke or cerebral infarct, Parkinson's disease

psychhx Psychiatric history, active psychosis or severe depression

chrpulhx Chronic pulmonary disease, severe pulmonary disease

renalhx Chronic renal disease, chronic hemodialysis or peritoneal dialysis

liverhx Cirrhosis, hepatic failure

gibledhx Upper GI bleeding

malighx Solid tumor, metastatic disease, chronic leukemia/myeloma, acute leukemia, lymphoma

immunhx Immunosuppression, organ transplant, HIV, Diabetes Mellitus, Connective Tissue Disease

transhx transfer (> 24 hours) from another hospital

amihx Definite myocardial infarction

das2d3pc DASI - Duke Activity Status Index

surv2md1 Estimate of prob. of surviving 2 months

aps1 APACHE score

scoma1 Glasgow coma score

wtkilo1 Weight

temp1 Temperature

meanbp1 Mean Blood Pressure

resp1 Respiratory Rate

hrt1 Heart Rate

pafi1 PaO2/FI02 ratio

paco21 PaCO2

ph1 PH

wblc1 WBC

hema1 Hematocrit

sod1 Sodium

pot1 Potassium

crea1 Creatinine

bili1 Bilirubin

alb1 Albumin

cat1_CHF 1 if the primary disease category is CHF, and 0 otherwise (Omitted category = ARF).

cat1_Cirrhosis 1 if the primary disease category is Cirrhosis, and 0 otherwise (Omitted category = ARF).

cat1_Colon_Cancer 1 if the primary disease category is Colon Cancer, and 0 otherwise (Omitted category = ARF).

cat1_Coma 1 if the primary disease category is Coma, and 0 otherwise (Omitted category = ARF).

cat1_COPD 1 if the primary disease category is COPD, and 0 otherwise (Omitted category = ARF).

cat1_Lung_Cancer 1 if the primary disease category is Lung Cancer, and 0 otherwise (Omitted category = ARF).

cat1_MOSF_Malignancy 1 if the primary disease category is MOSF w/Malignancy, and 0 otherwise (Omitted category = ARF).

cat1_MOSF_Sepsis 1 if the primary disease category is MOSF w/Sepsis, and 0 otherwise (Omitted category = ARF).

ca_Metastatic 1 if cancer is metastatic, and 0 otherwise (Omitted category = no cancer).

ca_Yes 1 if cancer is localized, and 0 otherwise (Omitted category = no cancer).

ninsclas_Medicaid 1 if medical insurance category is Medicaid, and 0 otherwise (Omitted category = Private).

ninsclas_Medicare 1 if medical insurance category is Medicare, and 0 otherwise (Omitted category = Private).

ninsclas_Medicare_and_Medicaid 1 if medical insurance category is Medicare & Medicaid, and 0 otherwise (Omitted category = Private).

ninsclas_No_insurance 1 if medical insurance category is No Insurance, and 0 otherwise (Omitted category = Private).

ninsclas_Private_and_Medicare 1 if medical insurance category is Private & Medicare, and 0 otherwise (Omitted category = Private).

race_black 1 if Black, and 0 otherwise (Omitted category = White).

race_other 1 if Other, and 0 otherwise (Omitted category = White).

income3 1 if Income >$50k, and 0 otherwise (Omitted category = under $11k).

income1 1 if Income $11–$25k, and 0 otherwise (Omitted category = under $11k).

income2 1 if Income $25–$50k, and 0 otherwise (Omitted category = under $11k).

resp_Yes Respiratory diagnosis

card_Yes Cardiovascular diagnosis

neuro_Yes Neurological diagnosis

gastr_Yes Gastrointestinal diagnosis

renal_Yes Renal diagnosis

meta_Yes Metabolic diagnosis

hema_Yes Hematological diagnosis

seps_Yes Sepsis diagnosis

trauma_Yes Trauma diagnosis

ortho_Yes Orthopedic diagnosis

dnr1_Yes Do Not Resuscitate status on day 1

sex_Female Female

cat2_Cirrhosis 1 if the secondary disease category is Cirrhosis, and 0 otherwise (Omitted category = NA).

cat2_Colon_Cancer 1 if secondary disease category is Colon Cancer, and 0 otherwise (Omitted category = NA).

cat2_Coma 1 if the secondary disease category is Coma, and 0 otherwise (Omitted category = NA).

cat2_Lung_Cancer 1 if the secondary disease category is Lung Cancer, and 0 otherwise (Omitted category = NA).

cat2_MOSF_Malignancy 1 if the secondary disease category is MOSF w/Malignancy, and 0 otherwise (Omitted category = NA).

cat2_MOSF_Sepsis 1 if the secondary disease category is MOSF w/Sepsis, and 0 otherwise (Omitted category = NA).

wt0 weight = 0 (missing)

Source

The dataset is publicly available on the Vanderbilt Biostatistics website at https://hbiostat.org/data/.

References

Connors, A.F., Speroff, T., Dawson, N.V., Thomas, C., Harrell, F.E., Wagner, D., Desbiens, N., Goldman, L., Wu, A.W., Califf, R.M. and Fulkerson, W.J., 1996. The effectiveness of right heart catheterization in the initial care of critically III patients. JAMA, 276(11), pp.889-897. doi: 10.1001/jama.1996.03540110043030

Hirano, K., Imbens, G.W. Estimation of Causal Effects using Propensity Score Weighting: An Application to Data on Right Heart Catheterization, 2001. Health Services & Outcomes Research Methodology 2, pp.259–278. doi: 10.1023/A:1020371312283

D. J. Murphy, L. E. Cluff, SUPPORT: Study to understand prognoses and preferences for outcomes and risks of treatments—study design, 1990. Journal of Clinical Epidemiology, 43, pp. 1S–123S https://www.jclinepi.com/issue/S0895-4356(00)X0189-8 .

`simulation_dgp` *Simulating observations from the data-generating process considered in Lee and Weidner (2021)*

Description

Simulates observations from the data-generating process considered in Lee and Weidner (2021)

Usage

```
simulation_dgp(n, ps_spec = "overlap", x_discrete = FALSE)
```

Arguments

`n`	sample size
`ps_spec`	specification of the propensity score: "overlap" or "non-overlap" (default: "overlap")
`x_discrete`	TRUE if the distribution of the covariate is uniform on -3.0, -2.9, ..., 3.0 and FALSE if the distribution of the covariate is uniform on [–3,3] (default: FALSE)

Value

An S3 object of type "ATbounds". The object has the following elements.

`outcome`	n observations of binary outcomes
`treat`	n observations of binary treatments
`covariate`	n observations of a scalar covariate
`ate_oracle`	the sample analog of E[Y(1) - Y(0)]
`att_oracle`	the sample analog of E[DY(1) - Y(0)|D=1]

References

Sokbae Lee and Martin Weidner. Bounding Treatment Effects by Pooling Limited Information across Observations.

Examples

```
data <- simulation_dgp(100, ps_spec = "overlap")
y <- data$outcome
d <- data$treat
x <- data$covariate
ate <- data$ate_oracle
att <- data$att_oracle
```

`summary.ATbounds` *Summary method for ATbounds objects*

Description

Produce a summary for an ATbounds object.

Usage

```
## S3 method for class 'ATbounds'
summary(object, ...)
```

Arguments

`object`	ATbounds object
`...`	Additional arguments for summary generic

Value

A summary is produced with bounds estimates and confidence intervals. In addition, it has the following elements.

`Lower_Bound`	lower bound estimate and lower end point of the confidence interval
`Upper_Bound`	upper bound estimate and upper end point of the confidence interval

References

Sokbae Lee and Martin Weidner. Bounding Treatment Effects by Pooling Limited Information across Observations.

Examples

```
Y <- RHC[,"survival"]
D <- RHC[,"RHC"]
X <- RHC[,c("age","edu")]
rps <- rep(mean(D),length(D))
results_ate <- atebounds(Y, D, X, rps, Q = 3)
summary(results_ate)
```

Index

∗ **datasets**
EFM, 5
RHC, 6

atebounds, 2
attbounds, 3

EFM, 5

RHC, 6

simulation_dgp, 9
summary.ATbounds, 10