

Package ‘exr’

July 16, 2022

Title Quantifying Robustness to External Validity Bias

Version 0.1.0

Author Naoki Egami and Martin Devaux

Maintainer Naoki Egami <naoki.egami@columbia.edu>

Description R package exr estimates a measure of external robustness (ranges from 0 to 1) by estimating how much different populations should be from the experimental sample to explain away the T-PATE. This package provides the central function `exr()` and companion functions `summary()` to summarize results and `plot()` to visualize results. Please read Devaux and Egami (2022+) (https://naokiegami.com/paper/external_robust.pdf).

License GPL-2

Encoding UTF-8

LazyData true

RoxygenNote 7.2.0

Depends R (>= 3.5.0)

Imports CVXR,
grf,
SuperLearner,
bartCause,
furr,
future,
future.apply,
progressr

Suggests rmarkdown,
knitr

VignetteBuilder knitr

R topics documented:

| | |
|-----------------------------|---|
| bench_mturk | 2 |
| bench_prob_survey | 2 |
| constraint | 2 |
| exr | 3 |
| plot.exr | 4 |
| summary.exr | 5 |

| | |
|-------|---|
| Index | 6 |
|-------|---|

| | |
|---|--|
| bench_mturk | <i>Benchmark Dataset based on Mturk Samples</i> |
| Description Benchmark Dataset based on Mturk Samples | |
| Usage bench_mturk | |
| Format data.frame | |
| bench_prob_survey | <i>Benchmark Dataset based on National Probability Surveys</i> |
| Description Benchmark Dataset based on National Probability Surveys | |
| Usage bench_prob_survey | |
| Format data.frame | |
| constraint | <i>Adding Constraints</i> |
| Description Adding Constraints | |
| Usage constraint(vars, type, values) | |
| Arguments | |
| vars | Names of variables that users want to add constraints. |
| type | One of the three types; "between", "less than or equal to", "larger than or equal to". |
| values | Values of the weighted mean of variables specified in 'vars'. When 'type = "between"', 'values' should be a vector of length 2. Otherwise, 'values' should be one numeric value. |

Description

Estimating External Robustness

Usage

```
exr(
  outcome,
  treatment,
  covariates,
  data,
  sate_estimate = NULL,
  family = "gaussian",
  threshold = 0,
  cate_method = "grf",
  cate_name = NULL,
  uncertainty = TRUE,
  ci = 0.95,
  boot = FALSE,
  nboot = 100,
  clusters = NULL,
  numCores = 1,
  const_list = NULL,
  lib = c("mean", "glm", "ranger"),
  verbose = FALSE,
  seed = 1234
)
```

Arguments

| | |
|---------------|---|
| outcome | Name of the outcome variable. |
| treatment | Name of the treatment variable. The treatment variable needs to be a binary variable. |
| covariates | Name of the covariates |
| data | 'data' needs to be 'data.frame' |
| sate_estimate | A vector of length 2; a point estimate of the SATE and its standard error. Default = NULL. When NULL, the package internally estimates the SATE using a linear regression of the outcome on the treatment and all specified covariates. |
| family | "gaussian" (continuous outcomes) or "binomial" (binary outcomes). Default = 'gaussian'. |
| threshold | Numeric. This is the threshold for the T-PATE. Default = 0 |
| cate_method | When "grf", we use 'grf'. When "X-learner", we use 'X-learner' based on R package 'SuperLearner'. |
| cate_name | Name of columns in 'data' that store CATEs estimated outside the function. Default = NULL, and the CATE is internally estimated by a method specified by 'cate_method.' |

| | |
|-------------|---|
| uncertainty | Logical. Whether we take into account uncertainties to estimate external robustness. Default is TRUE. |
| ci | Numeric. Specify the level of the confidence interval. Default = 0.95. Only used when 'uncertainty = TRUE'. |
| boot | Logical. When 'boot = TRUE', we use bootstrap to approximate confidence intervals. When 'boot = FALSE', we approximate the standard error using the standard error of the SATE (This is computationally much faster, so researchers can use this as an initial check. But for final results, we recommend 'boot = TRUE'). |
| nboot | Numeric. The number of bootstrap. Default = 100. Only used when 'uncertainty = TRUE' and 'boot = TRUE'. |
| clusters | Vector. Unique identifiers for computing cluster standard errors. Only used when 'uncertainty = TRUE' and 'boot = TRUE'. |
| numCores | Numeric. Default = 1. Number of cores to use. Parallel computing based on 'future' package is used only when 'uncertainty = TRUE' and 'boot = TRUE'. When 'numCores = NULL', it automatically detects the number of available cores. |
| const_list | List. Constraints to incorporate partial knowledge about population data. See Examples. |
| lib | (optional) The library used for 'SuperLearner' when we choose 'cate_method = "X-learner"'. Default = c("mean", "glm", "ranger"). |
| verbose | Logical. Whether to see outputs from the underlying optimization package 'CVXR'. Default = FALSE. |
| seed | Numeric. 'seed' used internally. Default = 1234. |

Value

exr returns an object of exr class.

- est: Estimated external robustness.
- est_unc: Estimates of the pAMCE for all factors in each bootstrap sample.
- tau: Estimated conditional average treatment effects (CATEs) for each unit.
- w: Estimated weights that solves the KL minimization problem where a constraint focuses on a point estimate of the T-PATE.
- w_unc: Estimated weights that solves the KL minimization problem where a constraint focuses on a confidence interval of the T-PATE.
- ...: Values for internal use.

| | |
|----------|--|
| plot.exr | <i>Visualize the distribution of the CATEs and estimated external robustness</i> |
|----------|--|

Description

Visualize the distribution of the CATEs and estimated external robustness

Usage

```
## S3 method for class 'exr'  
plot(x, cate_method = NULL, uncertainty = FALSE, ...)
```

Arguments

| | |
|-------------|---|
| x | the output from exr() |
| cate_method | When "grf", we use 'grf'. When "X-learner", we use 'X-learner' based on R package 'SuperLearner'. |
| uncertainty | Logical. Whether we visualize the distribution of the CATEs and estimated external robustness based on an adjusted threshold that takes into account uncertainties. |
| ... | Other arguments. |

summary.exr

Summary function

Description

Summary function

Usage

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

Arguments

| | |
|--------|-----------------------|
| object | the output from exr() |
| ... | Other arguments. |

Index

- * **dataset**
 - bench_mturk, [2](#)
 - bench_prob_survey, [2](#)
- bench_mturk, [2](#)
- bench_prob_survey, [2](#)
- constraint, [2](#)
- exr, [3](#)
- plot.exr, [4](#)
- summary.exr, [5](#)