

Package ‘psrwe’

September 8, 2020

Title PS-Integrated Methods for Incorporating RWE in Clinical Studies

Version 1.2

Author Chenguang Wang [aut, cre]

Trustees of Columbia University [cph] (tools/make_cpp.R, R/stanmodels.R)

Maintainer Chenguang Wang <cwang68@jhmi.edu>

Description High-quality real-world data can be transformed into scientific real-world evidence (RWE) for regulatory and healthcare decision-making using proven analytical methods and techniques. For example, propensity score (PS) methodology can be applied to pre-select a subset of real-world data containing patients that are similar to those in the current clinical study in terms of covariates, and to stratify the selected patients together with those in the current study into more homogeneous strata. Then, methods such as power prior approach or composite likelihood approach can be applied in each stratum to draw inference for the parameters of interest. This package provides functions that implement the PS-integrated RWE analysis methods proposed in Wang et al. (2019) <doi:10.1080/10543406.2019.1657133>, Wang et al. (2020) <doi:10.1080/10543406.2019.1684309> and Chen et al. (2020) <doi:10.1080/10543406.2020.1730877>.

Depends methods, R (>= 4.0), rstan (>= 2.19.3), Rcpp (>= 1.0.5)

License GPL (>= 3)

LinkingTo BH (>= 1.72.0-3), rstan (>= 2.19.3), Rcpp (>= 1.0.5), RcppEigen (>= 0.3.3.7.0), StanHeaders (>= 2.21.0-5)

RcppModules stan_fit4powerp_mod, stan_fit4powerps_mod, stan_fit4powerpsbinary_mod, stan_fit4prior_mod

Imports parallel (>= 3.2), cowplot (>= 1.0.0), dplyr (>= 0.8.5), ggplot2 (>= 3.3.2), randomForest (>= 4.6-14)

Suggests knitr, markdown

Encoding UTF-8

LazyData true

ByteCompile true

SystemRequirements GNU make

NeedsCompilation yes

RoxygenNote 7.1.0**VignetteBuilder** knitr**Repository** CRAN**Date/Publication** 2020-09-08 08:20:08 UTC**R topics documented:**

psrwe-package	2
ex_dta	3
ex_dta_rct	4
get_distance	4
plot.RWE_DWITHPS	5
rwe_cl	5
rwe_cut	6
rwe_ps	7
rwe_ps_borrow	8
rwe_ps_cl	9
rwe_ps_cl2arm	10
rwe_ps_dist	10
rwe_ps_match	11
rwe_ps_powerp	12
rwe_stan	13
summary.RWE_CLRST	14
summary.RWE_POWERPRST	15
Index	16

psrwe-package

*PS-Integrated Methods for Incorporating RWE in Clinical Studies***Description**

This package provide R functions for conducting clinical studies with real-world evidence incorporated in the study design and analysis.

PS-integrated power prior

We extend the Bayesian power prior approach for a single-arm study (the current study) to leverage external real-world data. We use propensity score methodology to pre-select a subset of real-world data containing patients that are similar to those in the current study in terms of covariates, and to stratify the selected patients together with those in the current study into more homogeneous strata. The power prior approach is then applied in each stratum to obtain stratum-specific posterior distributions, which are combined to complete the Bayesian inference for the parameters of interest.

PS-integrated composite likelihood

A propensity score-integrated composite likelihood (PSCL) approach is developed for cases in which the control arm of a two-arm randomized controlled trial (RCT) (treated vs control) is augmented with patients from real-world data (RWD) containing both clinical outcomes and covariates at the patient-level. The PSCL approach first estimates the propensity score for every patient as the probability of the patient being in the RCT rather than the RWD, and then stratifies all patients into strata based on the estimated propensity scores. Within each propensity score stratum, a composite likelihood function is specified and utilized to down-weight the information contributed by the RWD source. Estimates of the stratum-specific parameters are obtained by maximizing the composite likelihood function. These stratum-specific estimates are then combined to obtain an overall population-level estimate of the parameter of interest.

References

Chen, W.C., Wang, C., Li, H., Lu, N., Tiwari, R., Xu, Y. and Yue, L.Q., 2020. Propensity score-integrated composite likelihood approach for augmenting the control arm of a randomized controlled trial by incorporating real-world data. *Journal of Biopharmaceutical Statistics*, 30(3), pp.508-520.

Wang, C., Lu, N., Chen, W. C., Li, H., Tiwari, R., Xu, Y., & Yue, L. Q. (2020). Propensity score-integrated composite likelihood approach for incorporating real-world evidence in single-arm clinical studies. *Journal of biopharmaceutical statistics*, 30(3), 495-507.

Wang, C., Li, H., Chen, W. C., Lu, N., Tiwari, R., Xu, Y., & Yue, L. Q. (2019). Propensity score-integrated power prior approach for incorporating real-world evidence in single-arm clinical studies. *Journal of biopharmaceutical statistics*, 29(5), 731-748.

ex_dta

Example dataset

Description

Example dataset of a single arm study

Format

A dataframe with the following variables:

- Groupcurrent, rwd
- YBinary outcome
- V1-V7Covariates

 ex_dta_rct

Example dataset

Description

Example dataset of a randomized study

Format

A dataframe with the following variables:

- Group0,1
- Arm0, 1
- YContinuous outcome
- V1-V7Covariates

 get_distance

Distance between two distributions

Description

Get balance by different metrics

Usage

```
get_distance(cov0, cov1, metric = c("ovl", "ksd", "std", "abd", "ley", "mhb"))
```

Arguments

cov0	Vector (or matrix for mhb) of samples from the first distribution
cov1	Vector (or matrix for mhb) of samples from the second distribution
metric	Metrics of distances with options ovl Overlapping area ksd Kullback-Leibler distance std Standardized difference in mean abb Absolute difference in mean ley Levy distance mhb Mahalanobis distance

Value

A real value of the distance

Examples

```
x <- rnorm(100, mean = 0, sd = 1)
y <- rnorm(1000, mean = 1, sd = 2)
get_distance(x, y, "ovl")
get_distance(x, y, "abd")
```

plot.RWE_DWITHPs *Plot PS distributions*

Description

S3 method for visualizing PS adjustment

Usage

```
## S3 method for class 'RWE_DWITHPs'
plot(x, plot_type = c("ps", "balance"), ...)
```

Arguments

x	Class RWE_DWITHPs created by rwe_ps
plot_type	Types of plots. <ul style="list-style-type: none"> • psPS density plot • balanceCovariate balance plot
...	Additional parameter for the plot

rwe_cl *Composite Likelihood Estimation*

Description

Estimate parameter of interest based composite likelihood for a single PS stratum

Usage

```
rwe_cl(
  dta_cur,
  dta_ext,
  n_borrow = 0,
  outcome_type = c("continuous", "binary"),
  equal_sd = TRUE
)
```

Arguments

dta_cur	Vector of outcome from a PS stratum in current study
dta_ext	Vector of outcome from a PS stratum in external data source
n_borrow	Number of subjects to be borrowed
outcome_type	Type of outcomes: continuous or binary
equal_sd	Boolean. whether sd is the same between the current study and external data source

Value

Maximum composite likelihood estimator of the mean

Examples

```
x <- rnorm(100, mean = 0, sd = 1)
y <- rnorm(1000, mean = 1, sd = 2)
rwe_cl(x, y, n_borrow = 20, equal_sd = FALSE)
```

rwe_cut

Create strata

Description

The cut points are chosen such that there will with equal numbers in each bin for x . By default, values of y that are outside the range of x will be excluded from the bins, unless they are in the `keep_inx`.

Usage

```
rwe_cut(x, y = x, breaks = 5, keep_inx = NULL)
```

Arguments

x	Vector of values based on which cut points will be determined
y	Vector of values to be cut, default to be the same as x
breaks	Number of cut points
keep_inx	Indices of y that will be categorized as 1 or the largest bin even if their values are out of range of x , i.e. the y 's that will not be trimmed

Details

Cut a sequence of numbers into bins.

Value

A vector of stratum assignment for y . The y 's that are outside the range of x and not in `keep_inx` are assigned NA in the result.

Examples

```
x <- rnorm(100, mean = 0, sd = 1)
y <- rnorm(1000, mean = 1, sd = 2)
rwe_cut(x, y, breaks = 5)
```

rwe_ps	<i>Get propensity scores</i>
--------	------------------------------

Description

Calculate propensity scores using logistic regression or random forest model

Usage

```
rwe_ps(
  data,
  ps_fm1 = NULL,
  v_covs = "V1",
  v_grp = "Group",
  cur_grp_level = 1,
  v_arm = NULL,
  trt_arm_level = 1,
  nstrata = 5,
  ...
)
```

Arguments

<code>data</code>	Dataframe with group assignment and covariates
<code>ps_fm1</code>	Propensity score formula. If null, all covariates will be included in the PS model in a linear form.
<code>v_covs</code>	Vector of column names corresponding to covariates
<code>v_grp</code>	Column name corresponding to group assignment
<code>cur_grp_level</code>	Group level for the current study. Default 1.
<code>v_arm</code>	Column name corresponding treatment vs. control. Ignored for single arm studies. two-arm randomized studies.
<code>trt_arm_level</code>	Arm level for the treatment arm. Ignored for single-arm studies.
<code>nstrata</code>	Number of PS strata to be created

... Parameters to get propensity scores.

- ps_method `logistic` for logistic regression or `randomforest` for randomforest) ...Parameters for `randomForest`

Value

A class `RWE_DWITHPS` list with items:

- data Original data with column `_ps_` for estimated PS scores and `_strata_` for PS stratum added
- ps_fm1PS formula for estimated PS scores
- nstrata Number of strata

Examples

```
rwe_ps(ex_dta, v_covs = paste("V", 1:7, sep = ""), v_grp = "Group",
       cur_grp_level = "current")
```

<code>rwe_ps_borrow</code>	<i>Get number of subjects borrowed from each stratum</i>
----------------------------	--

Description

Based on PS distances, split the total number of subjects to be borrowed from the external data source to each stratum

Usage

```
rwe_ps_borrow(total_borrow, dta_ps_dist, ...)
```

Arguments

<code>total_borrow</code>	Total number of subjects to be borrowed
<code>dta_ps_dist</code>	A class <code>RWE_PSDIST</code>
...	Other options

Value

A vector of number of subjects to be borrowed from each stratum

Examples

```
dta_ps <- rwe_ps(ex_dta,
  v_covs = paste("V", 1:7, sep = ""),
  v_grp = "Group",
  cur_grp_level = "current")

ps_dist <- rwe_ps_dist(dta_ps, metric = "ovl")
ps_borrow <- rwe_ps_borrow(total_borrow = 100, ps_dist)
```

rwe_ps_cl

*PS-Integrated Composite Likelihood Estimation***Description**

Estimate the mean of the outcome based on PS-integrated composite likelihood approach. Variance is estimated by Jack-Knife method. Applies to the case when there is only one external data source.

Usage

```
rwe_ps_cl(data_withps, v_borrow = 0, v_outcome = "Y", ...)
```

Arguments

data_withps	A class RWE_DWITHPS list. See rwe_ps .
v_borrow	Vector of number of subjects to be borrowed from each stratum
v_outcome	Column name corresponding to the outcome
...	Parameters for rwe_cl

Value

A dataframe with class name RWE_CLRST. It contains the composite estimation of the mean for each stratum as well as the jackknife estimation for each subject. The results should be further summarized by its S3 method `summary`.

Examples

```
dta_ps <- rwe_ps(ex_dta,
  v_covs = paste("V", 1:7, sep = ""),
  v_grp = "Group",
  cur_grp_level = "current",
  nstrata = 5)
ps_dist <- rwe_ps_dist(dta_ps)
ps_borrow <- rwe_ps_borrow(total_borrow = 40, ps_dist)
rst_cl <- rwe_ps_cl(dta_ps, v_borrow = ps_borrow)
summary(rst_cl)
```

rwe_ps_cl2arm	<i>PS-Integrated Composite Likelihood Estimation for Randomized Study</i>
---------------	---

Description

Estimate the treatment effect based on PS-integrated composite likelihood approach. Subjects from the external data source augment an arm, usually the control arm.

Usage

```
rwe_ps_cl2arm(
  data_withps,
  v_arm = "Arm",
  trt_arm_level = 1,
  total_borrow = 0,
  ...
)
```

Arguments

data_withps	A class RWE_DWITHPS list. See rwe_ps .
v_arm	Column name corresponding treatment vs. control. Ignored for single arm studies. two-arm randomized studies.
trt_arm_level	Arm level for the treatment arm. Ignored for single-arm studies.
total_borrow	Total number of subjects to be borrowed
...	Parameters for rwe_ps_cl

Value

A list with estimated treatment effect and the variance, as well as the estimated mean for the treatment and control arms

rwe_ps_dist	<i>Summarize PS scores</i>
-------------	----------------------------

Description

Get number of subjects and the distances of PS distributions for each PS strata

Usage

```
rwe_ps_dist(data_withps, min_n0 = 10, trt_arm_level = NULL, ...)
```

Arguments

data_withps	A class RWE_DWITHPS list. See rwe_ps .
min_n0	threshold for number of external subjects, below which the external data in the current stratum will be ignored by setting the PS distance to 0. Default value 10.
trt_arm_level	Arm level for the treatment arm. Ignored for single-arm studies.
...	Parameters for <code>get_distance</code> , e.g., <code>metric</code> with options such as <code>overlapping area (ovl)</code> .

Value

A class RWE_PSDIST dataframe with columns

- StrataIndex of stratum. 0 represents the overall information
- N0,N1Number of subjects in group 0 and 1
- N00, N10Number of arm 0 subjects in group 0 and 1, when arm exists
- N01, N11Number of arm 1 subjects in group 0 and 1, when arm exists
- Dist Distance

Examples

```
dta_ps <- rwe_ps(ex_dta,
  v_covs = paste("V", 1:7, sep = ""),
  v_grp = "Group",
  cur_grp_level = "current")

rwe_ps_dist(dta_ps, metric = "ovl")
```

rwe_ps_match

PS matching

Description

Match patients in external data source with patients in current study based on PS using nearest neighbor method

Usage

```
rwe_ps_match(dta_cur, dta_ext, ratio = 3, ps_fm1 = NULL, v_covs = "V1")
```

Arguments

dta_cur	current study dataframe
dta_ext	external data source dataframe
ratio	matching ratio
ps_fml	Propensity score formula. If null, all covariates will be included in the PS model in a linear form.
v_covs	Vector of column names corresponding to covariates

Value

A matrix with two columns. Column pid: id in dta_cur; Column match_id: matched id from dta_ext.

Examples

```
dta_cur <- ex_dta[which(ex_dta$Group == "current"), ]
dta_ext <- ex_dta[which(ex_dta$Group != "current"), ]
rwe_ps_match(dta_cur, dta_ext, v_covs = paste("V", 1:7, sep = ""),
             ratio = 2)
```

rwe_ps_powerp

Get posterior samples based on PS-power prior approach

Description

Draw posterior samples of the parameters of interest for the PS-power prior approach

Usage

```
rwe_ps_powerp(
  data_withps,
  outcome_type = c("continuous", "binary"),
  total_borrow = 0,
  v_distance = NULL,
  prior_type = c("fixed", "random"),
  v_outcome = "Y",
  ...,
  seed = NULL
)
```

Arguments

<code>data_withps</code>	A class RWE_DWITHPS list. See rwe_ps .
<code>outcome_type</code>	Type of outcomes: continuous or binary
<code>total_borrow</code>	Total number of subjects to be borrowed
<code>v_distance</code>	Vector of distances in PS distributions for each stratum. See rwe_ps_dist .
<code>prior_type</code>	Whether treat power parameter as fixed (<code>fixed</code>) or fully Bayesian (<code>random</code>)
<code>v_outcome</code>	Column name corresponding to the outcome
<code>...</code>	extra parameters for calling function rwe_stan
<code>seed</code>	Random seed

Value

A list with the following objects

post_theta Posterior samples of parameter of interest

post_theta_stratum Posterior samples of parameter of interest in each stratum

stan_rst Result from STAN sampling

Examples

```
dta_ps <- rwe_ps(ex_dta, v_covs = paste("V", 1:10, sep = ""),
                v_grp = "group", cur_grp_level = "current")
ps_dist <- rwe_ps_dist(dta_ps, metric = "ov1")
post_smps <- rwe_ps_powerp(dta_ps, outcome_type = "binary",
                          total_borrow = 30,
                          v_distance = ps_dist$Dist[1:5],
                          v_outcome = "Y")
```

rwe_stan

Call STAN models

Description

Call STAN models. Called by `rwe_ps_powerp`.

Usage

```
rwe_stan(
  lst_data,
  stan_mdl = c("powerps", "powerpsbinary", "powerp"),
  chains = 4,
  iter = 2000,
  warmup = 1000,
  control = list(adapt_delta = 0.95),
  ...
)
```

Arguments

lst_data	List of study data to be passed to STAN
stan_md1	STAN model including powerps PS-power prior model for continuous outcomes powerpsbinary PS-power prior model for binary outcomes powerp Power prior model
chains	STAN parameter. Number of Markov chainsm
iter	STAN parameter. Number of iterations
warmup	STAN parameter. Number of burnin.
control	STAN parameter. See rstan::stan for details.
...	other options to call STAN sampling such as thin, algorithm. See rstan::sampling for details.#'

Value

Result from STAN sampling

summary.RWE_CLRST	<i>Summary of Composite Likelihood Estimation</i>
-------------------	---

Description

Summarize the result in a class RWE_CLRST objects generated by [rwe_ps_cl](#).

Usage

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

Arguments

object	A class RWE_CLRST object
...	Extra arguments

Value

Summary of composite likelihood estimation

summary.RWE_POWERPRST *Summary of PS-power prior Estimation*

Description

Summarize the result in a class RWE_POWERPRST objects generated by [rwe_ps_cl](#).

Usage

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

Arguments

object	A class RWE_POWERPRST object
...	Extra arguments

Value

Summary of composite likelihood estimation

Index

`ex_dta`, [3](#)
`ex_dta_rct`, [4](#)

`get_distance`, [4](#)

`plot.RWE_DWITHPS`, [5](#)
`psrwe` (`psrwe`-package), [2](#)
`psrwe`-package, [2](#)

`rwe_cl`, [5](#)
`rwe_cut`, [6](#)
`rwe_ps`, [5](#), [7](#), [9–11](#), [13](#)
`rwe_ps_borrow`, [8](#)
`rwe_ps_cl`, [9](#), [10](#), [14](#), [15](#)
`rwe_ps_cl2arm`, [10](#)
`rwe_ps_dist`, [10](#), [13](#)
`rwe_ps_match`, [11](#)
`rwe_ps_powerp`, [12](#)
`rwe_stan`, [13](#), [13](#)

`summary.RWE_CLRST`, [14](#)
`summary.RWE_POWERPRST`, [15](#)