Package 'psBayesborrow'

July 23, 2025

Version 1.1.0

Title Bayesian Information Borrowing with Propensity Score Matching

Description Hybrid control design is a way to borrow information from external controls to augment concurrent controls in a randomized controlled trial and is expected to overcome the feasibility issue when adequate randomized controlled trials cannot be conducted. A major challenge in the hybrid control design is its inability to eliminate a prior-data conflict caused by systematic imbalances in measured or unmeasured confounding factors between patients in the concurrent treatment/control group and external controls. To prevent the prior-data conflict, a combined use of propensity score matching and Bayesian commensurate prior has been proposed in the context of hybrid control design. The propensity score matching is first performed to guarantee the balance in baseline characteristics, and then the Bayesian commensurate prior is constructed while discounting the information based on the similarity in outcomes between the concurrent and external controls. 'psBayesborrow' is a package to implement the propensity score matching and the Bayesian analysis with commensurate prior, as well as to conduct a simulation study to assess operating characteristics of the hybrid control design, where users can choose design parameters in flexible and straightforward ways depending on their own application.

License MIT + file LICENSE

Encoding UTF-8

RoxygenNote 7.3.1

Biarch true

Depends R (>= 3.4.0)

Imports methods, copula, Rcpp (>= 0.12.0), RcppParallel (>= 5.0.1), rstan (>= 2.18.1), rstantools (>= 2.2.0), boot, MatchIt, optmatch, survival, e1071, overlapping, stats

Suggests testthat (>= 3.0.0)

LinkingTo BH (>= 1.66.0), Rcpp (>= 0.12.0), RcppEigen (>= 0.3.3.3.0), RcppParallel (>= 5.0.1), rstan (>= 2.18.1), StanHeaders (>= 2.18.0)

SystemRequirements GNU make

NeedsCompilation yes

Author Yusuke Yamaguchi [aut, cre],
     Jun Takeda [aut],
     Kentaro Takeda [aut]

Maintainer Yusuke Yamaguchi <yamagubed@gmail.com>

Repository CRAN

Date/Publication 2024-03-04 22:30:02 UTC

Contents

---

psBayesborrow-package   *The 'psBayesborrow' package.*

Description

A DESCRIPTION OF THE PACKAGE

Author(s)

Maintainer: Yusuke Yamaguchi <yamagubed@gmail.com>

Authors:

- Jun Takeda

- Kentaro Takeda

References

Stan Development Team (NA). RStan: the R interface to Stan. R package version 2.26.13. https://mc-stan.org

---

| commensurate.bin | *Bayesian analysis with commensurate prior for binary outcome* |

Description

Bayesian analysis for binary outcome is implemented via MCMC, where a commensurate prior is used for incorporating data from external controls. No borrowing and full borrowing are also applicable.

Usage

```
commensurate.bin(
  formula, data, method.borrow,
  chains=2, iter=4000, warmup=floor(iter/2), thin=1,
  alternative="greater", sig.level=0.025,
  seed=sample.int(.Machine$integer.max,1))
```

Arguments

formula	Object of class formula, which is a symbolic description of the model to be fitted. The explanatory variables only include covariates of interest, which must be specified in the form of linear combination.
data	Data frame, which must have variables named study for study indicator (0 for external control, and 1 for current trial) and treat for treatment indicator (0 for concurrent and external control, and 1 for treatment).
method.borrow	List of information borrowing method. "noborrow" uses the concurrent data only. "fullborrow" uses the external control data without discounting. "cauchy" uses the commensurate prior to dynamically borrow the external control data, and the commensurability parameter is assumed to follow a half-Cauchy distribution. "normal" uses the commensurate prior to dynamically borrow the external control data, and the commensurability parameter is assumed to follow a half-normal distribution. "cauchy" and "normal" require to specify the scale parameter scale of half-Cauchy and half-normal distribution respectively.
chains	Number of Markov chains in MCMC sampling. The default value is chains=2.
iter	Number of iterations for each chain (including warmup) in MCMC sampling. The default value is iter=4000.
warmup	Number of warmup (aka burnin) iterations per chain in MCMC sampling. The default value is warmup=floor(iter/2).
thin	Period for saving samples in MCMC sampling. The default value is thin=1.
alternative	Alternative hypothesis to be tested ("greater" or "less"). The default value is alternative="greater".
sig.level	Significance level. The default value is sig.level=0.025.
seed	Setting a seed for MCMC sampling.

Details

The binary outcome is assumed to follow a binomial distribution. Given more than one covariates, a logistic regression model is built and its Bayesian estimation is performed via MCMC. Commensurate prior is used to dynamically discount the information to be borrowed from external control based on the similarity between the current trial and external controls, where the commensurability parameter determines the extent of borrowing. The commensurability parameter is assumed to follow a half-cauchy or a half-normal distribution, and its scale parameter needs to be carefully specified. No borrowing approach is to perform the analysis without incorporating the external controls. Full borrowing approach is just to pool the concurrent and external controls, which is used as a comparator in the analysis.

Value

The `commensurate.cont` returns a list containing the following objects:

reject	Data frame containing results of Bayesian one-sided hypothesis testing (whether or not the posterior probability that the log odds ratio is greater or less than 0 exceeds 1 minus significance level): TRUE when significant, otherwise FALSE.
theta	Data frame containing posterior mean, median, and sd of log odds ratio.
stan.obj	Stanfit object.

References

Hobbs BP, Carlin BP, Mandrekar SJ, Sargent DJ. Hierarchical commensurate and power prior models for adaptive incorporation of historical information in clinical trials. *Biometrics* 2011; 67:1047-1056.

Hobbs BP, Sargent DJ, Carlin BP. Commensurate priors for incorporating historical information in clinical trials using general and generalized linear models. *Bayesian Analysis* 2012; 7:639-674.

Examples

```
n.CT  <- 100
n.CC  <- 50
n.ECp <- 200

out.prob.CT <- 0.2
out.prob.CC <- 0.2
driftOR     <- 1.0

cov.C <- list(list(dist="norm",mean=0,sd=1,lab="cov1"),
              list(dist="binom",prob=0.4,lab="cov2"))

cov.cor.C <- rbind(c(  1,0.1),
                   c(0.1,  1))

cov.EC <- list(list(dist="norm",mean=0,sd=1,lab="cov1"),
               list(dist="binom",prob=0.4,lab="cov2"))

cov.cor.EC <- rbind(c(  1,0.1),
                    c(0.1,  1))
```

```
cov.effect <- c(0.9,0.9)

indata <- trial.simulation.bin(
  n.CT=n.CT, n.CC=n.CC, n.ECp=n.ECp,
  out.prob.CT=out.prob.CT, out.prob.CC=out.prob.CC, driftOR=driftOR,
  cov.C=cov.C, cov.cor.C=cov.cor.C,
  cov.EC=cov.EC, cov.cor.EC=cov.cor.EC, cov.effect=cov.effect)

n.EC <- 50

method.whomatch <- "conc.treat"
method.matching <- "optimal"
method.psorder  <- NULL

out.psmatch <- psmatch(
  study~cov1+cov2, data=indata, n.EC=n.EC,
  method.whomatch=method.whomatch, method.matching=method.matching,
  method.psorder=method.psorder)

indata.match <- rbind(indata[indata$study==1,],indata[out.psmatch$subjid.EC,])

method.borrow <- list(list(prior="cauchy",scale=2.0),
                      list(prior="normal",scale=0.5))

commensurate.bin(y~cov1,data=indata.match,method.borrow=method.borrow,chains=1,iter=100)
```

---

commensurate.cont   *Bayesian analysis with commensurate prior for continuous outcome*

Description

Bayesian analysis for continuous outcome is implemented via MCMC, where a commensurate prior is used for incorporating data from external controls. No borrowing and full borrowing are also applicable.

Usage

```
commensurate.cont(
  formula, data, method.borrow,
  chains=2, iter=4000, warmup=floor(iter/2), thin=1,
  alternative="greater", sig.level=0.025,
  seed=sample.int(.Machine$integer.max,1))
```

Arguments

formula   Object of class formula, which is a symbolic description of the model to be
                 fitted. The explanatory variables only include covariates of interest, which must
                 be specified in the form of linear combination.

data   Data frame, which must have variables named study for study indicator (0 for
                     external control, and 1 for current trial) and treat for treatment indicator (0 for
                     concurrent and external control, and 1 for treatment).

method.borrow   List of information borrowing method. "noborrow" uses the concurrent data
                     only. "fullborrow" uses the external control data without discounting. "cauchy"
                     uses the commensurate prior to dynamically borrow the external control data,
                     and the commensurability parameter is assumed to follow a half-Cauchy dis-
                     tribution. "normal" uses the commensurate prior to dynamically borrow the
                     external control data, and the commensurability parameter is assumed to follow
                     a half-normal distribution. "cauchy" and "normal" require to specify the scale
                     parameter scale of half-Cauchy and half-normal distribution respectively.

chains   Number of Markov chains in MCMC sampling. The default value is chains=2.

iter   Number of iterations for each chain (including warmup) in MCMC sampling.
                     The default value is iter=4000.

warmup   Number of warmup (aka burnin) iterations per chain in MCMC sampling. The
                     default value is warmup=floor(iter/2).

thin   Period for saving samples in MCMC sampling. The default value is thin=1.

alternative   Alternative hypothesis to be tested ("greater" or "less"). The default value is
                     alternative="greater".

sig.level   Significance level. The default value is sig.level=0.025.

seed   Setting a seed for MCMC sampling.

Details

The continuous outcome is assumed to follow a normal distribution. Given more than one covari-
ates, a normal linear regression model is built and its Bayesian estimation is performed via MCMC.
Commensurate prior is used to dynamically discount the information to be borrowed from external
control based on the similarity between the current trial and external controls, where the commensu-
rability parameter determines the extent of borrowing. The commensurability parameter is assumed
to follow a half-cauchy or a half-normal distribution, and its scale parameter needs to be carefully
specified. No borrowing approach is to perform the analysis without incorporating the external con-
trols. Full borrowing approach is just to pool the concurrent and external controls, which is used as
a comparator in the analysis.

Value

The commensurate.cont returns a list containing the following objects:

reject   Data frame containing results of Bayesian one-sided hypothesis testing (whether
                     or not the posterior probability that the mean difference is greater or less than 0
                     exceeds 1 minus significance level): TRUE when significant, otherwise FALSE.

theta   Data frame containing posterior mean, median, and sd of mean difference.

stan.obj   Stanfit object.

References

Hobbs BP, Carlin BP, Mandrekar SJ, Sargent DJ. Hierarchical commensurate and power prior models for adaptive incorporation of historical information in clinical trials. *Biometrics* 2011; 67:1047-1056.

Hobbs BP, Sargent DJ, Carlin BP. Commensurate priors for incorporating historical information in clinical trials using general and generalized linear models. *Bayesian Analysis* 2012; 7:639-674.

Examples

```
n.CT  <- 100
n.CC  <- 50
n.ECp <- 200

out.mean.CT <- 0
out.sd.CT   <- 1
out.mean.CC <- 0
out.sd.CC   <- 1
driftdiff   <- 0
out.sd.EC   <- 1

cov.C <- list(list(dist="norm",mean=0,sd=1,lab="cov1"),
              list(dist="binom",prob=0.4,lab="cov2"))

cov.cor.C <- rbind(c(  1,0.1),
                   c(0.1,  1))

cov.EC <- list(list(dist="norm",mean=0,sd=1,lab="cov1"),
               list(dist="binom",prob=0.4,lab="cov2"))

cov.cor.EC <- rbind(c(  1,0.1),
                    c(0.1,  1))

cov.effect <- c(0.1,0.1)

indata <- trial.simulation.cont(
  n.CT=n.CT, n.CC=n.CC, n.ECp=n.ECp,
  out.mean.CT=out.mean.CT, out.sd.CT=out.sd.CT,
  out.mean.CC=out.mean.CC, out.sd.CC=out.sd.CC,
  driftdiff=driftdiff, out.sd.EC=out.sd.EC,
  cov.C=cov.C, cov.cor.C=cov.cor.C,
  cov.EC=cov.EC, cov.cor.EC=cov.cor.EC, cov.effect=cov.effect)

n.EC <- 50

method.whomatch <- "conc.treat"
method.matching <- "optimal"
method.psorder  <- NULL

out.psmatch <- psmatch(
  study~cov1+cov2, data=indata, n.EC=n.EC,
  method.whomatch=method.whomatch, method.matching=method.matching,
```

```
  method.psorder=method.psorder)

indata.match <- rbind(indata[indata$study==1,],indata[out.psmatch$subjid.EC,])

method.borrow <- list(list(prior="cauchy",scale=2.0),
                      list(prior="normal",scale=0.5))

commensurate.cont(y~cov1,data=indata.match,method.borrow=method.borrow,chains=1,iter=100)
```

---

commensurate.t2e	*Bayesian analysis with commensurate prior for time-to-event outcome*

Description

Bayesian analysis for time-to-event outcome is implemented via MCMC, where a commensurate prior is used for incorporating data from external controls. No borrowing and full borrowing are also applicable.

Usage

```
commensurate.t2e(
  formula, data, method.borrow,
  chains=2, iter=4000, warmup=floor(iter/2), thin=1,
  alternative="greater", sig.level=0.025,
  seed=sample.int(.Machine$integer.max,1))
```

Arguments

formula	Object of class formula, which is a symbolic description of the model to be fitted. The dependent variable must be an object of class Surv. The explanatory variables only include covariates of interest, which must be specified in the form of linear combination.
data	Data frame, which must have variables named study for study indicator (0 for external control, and 1 for current trial) and treat for treatment indicator (0 for concurrent and external control, and 1 for treatment).
method.borrow	List of information borrowing method. "noborrow" uses the concurrent data only. "fullborrow" uses the external control data without discounting. "cauchy" uses the commensurate prior to dynamically borrow the external control data, and the commensurability parameter is assumed to follow a half-Cauchy distribution. "normal" uses the commensurate prior to dynamically borrow the external control data, and the commensurability parameter is assumed to follow a half-normal distribution. "cauchy" and "normal" require to specify the scale parameter scale of half-Cauchy and half-normal distribution respectively.
chains	Number of Markov chains in MCMC sampling. The default value is chains=2.
iter	Number of iterations for each chain (including warmup) in MCMC sampling. The default value is iter=4000.

warmup	Number of warmup (burnin) iterations per chain in MCMC sampling. The default value is `warmup=floor(iter/2)`.
thin	Period for saving samples in MCMC sampling. The default value is `thin=1`.
alternative	Alternative hypothesis to be tested ("greater" or "less"). The default value is `alternative="greater"`.
sig.level	Significance level. The default value is `sig.level=0.025`.
seed	Setting a seed for MCMC sampling.

Details

The time to event outcome is assumed to follow a Weibull distribution. Given more than one covariates, a Weibull proportional hazards model is built and its Bayesian estimation is performed via MCMC. Commensurate prior is used to dynamically discount the information to be borrowed from external control based on the similarity between the current trial and external controls, where the commensurability parameter determines the extent of borrowing. The commensurability parameter is assumed to follow a half-cauchy or a half-normal distribution, and its scale parameter needs to be carefully specified. No borrowing approach is to perform the analysis without incorporating the external controls. Full borrowing approach is just to pool the concurrent and external controls, which is used as a comparator in the analysis.

Value

The `commensurate.t2e` returns a list containing the following objects:

reject	Data frame containing results of Bayesian one-sided hypothesis testing (whether or not the posterior probability that the log hazard ratio is greater or less than 0 exceeds 1 minus significance level): `TRUE` when significant, otherwise `FALSE`.
theta	Data frame containing posterior mean, median, and sd of log hazard ratio.
stan.obj	Stanfit object.

References

Hobbs BP, Carlin BP, Mandrekar SJ, Sargent DJ. Hierarchical commensurate and power prior models for adaptive incorporation of historical information in clinical trials. *Biometrics* 2011; 67:1047-1056.

Hobbs BP, Sargent DJ, Carlin BP. Commensurate priors for incorporating historical information in clinical trials using general and generalized linear models. *Bayesian Analysis* 2012; 7:639-674.

Examples

```
n.CT       <- 100
n.CC       <- 50
nevent.C   <- 100
n.ECp      <- 200
nevent.ECp <- 180
accrual    <- 16

out.mevent.CT <- 6
```

```
out.mevent.CC <- 6
driftHR       <- 1

cov.C <- list(list(dist="norm",mean=0,sd=1,lab="cov1"),
              list(dist="binom",prob=0.4,lab="cov2"))

cov.cor.C <- rbind(c(  1,0.1),
                   c(0.1,  1))

cov.EC <- list(list(dist="norm",mean=0,sd=1,lab="cov1"),
               list(dist="binom",prob=0.4,lab="cov2"))

cov.cor.EC <- rbind(c(  1,0.1),
                    c(0.1,  1))

cov.effect <- c(0.9,0.9)

indata <- trial.simulation.t2e(
  n.CT=n.CT, n.CC=n.CC, nevent.C=nevent.C,
  n.ECp=n.ECp, nevent.ECp=nevent.ECp, accrual=accrual,
  out.mevent.CT, out.mevent.CC, driftHR,
  cov.C=cov.C, cov.cor.C=cov.cor.C,
  cov.EC=cov.EC, cov.cor.EC=cov.cor.EC, cov.effect=cov.effect)

n.EC <- 50

method.whomatch <- "conc.treat"
method.matching <- "optimal"
method.psorder  <- NULL

out.psmatch <- psmatch(
  study~cov1+cov2, data=indata, n.EC=n.EC,
  method.whomatch=method.whomatch, method.matching=method.matching,
  method.psorder=method.psorder)

indata.match <- rbind(indata[indata$study==1,],indata[out.psmatch$subjid.EC,])

method.borrow <- list(list(prior="cauchy",scale=2.0),
                      list(prior="normal",scale=0.5))

commensurate.t2e(
  survival::Surv(time,status)~cov1+cov2,data=indata.match,
  method.borrow=method.borrow,chains=1,iter=100)
```

---

datagen	*Generate multivariate correlated data*

Description

Multivariate correlated data are generated. Gaussian copula is used to specify the correlation between variables. Any probability distributions available in R STAT is applicable.

Usage

```
datagen(margdist, corvec, nsim)
```

Arguments

margdist	List of distributions to be used for the data generation.
corvec	Vector of Gaussian copula correlation parameter.
nsim	Number of simulation.

Value

Data drawn from a multivariate distribution specified.

---

psborrow.bin	*Simulation study of hybrid control design with Bayesian dynamic borrowing incorporating propensity score matched external control: binary outcome*

Description

Simulation study is conducted to assess operating characteristics of hybrid control design with Bayesian dynamic borrowing, where the concurrent control is augmented by external control. The external controls are selected from external control pool using a propensity score matching. Commensurate power prior is used for Bayesian dynamic borrowing. The binary outcome is applicable.

Usage

```
psborrow.bin(
  n.CT, n.CC, n.ECp, n.EC,
  out.prob.CT, out.prob.CC, driftOR,
  cov.C, cov.cor.C, cov.EC, cov.cor.EC, cov.effect,
  psmatch.cov,
  method.psest="glm", method.pslink="logit",
  method.whomatch, method.matching, method.psorder, n.boot=100,
  analysis.cov, method.borrow,
  chains=2, iter=4000, warmup=floor(iter/2), thin=1,
  alternative="greater", sig.level=0.025, nsim)
```

Arguments

n.CT	Number of patients in treatment group in the current trial.
n.CC	Number of patients in concurrent control group in the current trial.
n.ECp	Number of patients in external control pool.
n.EC	Number of patients in external control.
out.prob.CT	True rate of outcome in treatment group in the current trial.

out.prob.CC	True rate of outcome in concurrent control group in the current trial.
driftOR	Odds ratio between concurrent and external control for which the bias should be plotted (odds in external control divided by odds in concurrent control).
cov.C	List of covariate distributions for treatment and concurrent control group in the current trial. Continuous and binary covariate are applicable. The continuous covariate is assumed to follow a normal distribution; for example, specified as list(dist="norm", mean=0, sd=1, lab="cov1"). The binary covariate is assumed to follow a binomial distribution; for example, specified as list(dist="binom", prob=0.4, lab="cov2"). lab is the column name of the covariate in the data frame generated.
cov.cor.C	Matrix of correlation coefficients for each pair of covariate for treatment and concurrent control group in the current trial, specified as Gaussian copula parameter.
cov.EC	List of covariate distributions for external control. The continuous covariate is assumed to follow a normal distribution; for example, specified as list(dist="norm", mean=0, sd=1, lab="cov1"). The binary covariate is assumed to follow a binomial distribution; for example, specified as list(dist="binom", prob=0.4, lab="cov2"). lab is the column name of the covariate in the data frame generated, which must be consistent with those used for cov.C.
cov.cor.EC	Matrix of correlation coefficients for each pair of covariate for external control, specified as Gaussian copula parameter.
cov.effect	Vector of covariate effects on the outcome, specified as odds ratio per one unit increase in continuous covariates or as odds ratio between categories for binary covariates.
psmatch.cov	Vector of names of covariates which are used for the propensity score matching. The names of covariates must be included in lab values specified in cov.C.
method.psest	Method of estimating the propensity score. Allowable options include, for example, "glm" for generalized linear model (e.g., logistic regression); "gam" for generalized additive model; "gbm" for generalized boosted model; "lasso" for lasso regression; "rpart" for classification tree. The default value is method.psest="glm".
method.pslink	Link function used in estimating the propensity score. Allowable options depend on the specific method.psest value specified. The default value is method.pslink="logit", which, along with method.psest="glm", identifies the default method as logistic regression.
method.whomatch	
	Options of who to match. Allowable options include conc.contl for matching concurrent control to external control pool; conc.treat for matching treatment to external control pool; conc.all for matching treatment plus concurrent control to external control pool; treat2contl for matching treatment to concurrent control plus external control pool.
method.matching	
	Matching method. Allowable options include "optimal" for optimal matching; "nearest" for nearest neighbor matching without replacement; "med.optimal" for equally splitting patients in the current trial and taking the median of each subset, followed by 1:1 optimal matching; "med.nearest" for equally splitting

patients in the current trial and taking the median of each subset, followed by 1:1 nearest neighbor matching without replacement; "km.optimal" for k-means clustering of patients in the current trial, followed by 1:1 optimal matching; "km.nearest" for k-means clustering of patients in the current trial, followed by 1:1 nearest neighbor matching without replacement; "cm.optimal" for fuzzy c-means clustering of patients in the current trial, followed by 1:1 optimal matching; "cm.nearest" for fuzzy c-means of patients in the current trial, followed by 1:1 nearest neighbor matching without replacement; "boot.optimal" for bootstrap sampling from patients in the current trial, followed by 1:1 optimal matching; "boot.nearest" for bootstrap sampling from patient in the current trial, followed by 1:1 nearest neighbor matching without replacement.

method.psorder	Order that the matching takes place when a nearest neighbor matching is used. Allowable options include "largest", where matching takes place in descending order of propensity score; "smallest", where matching takes place in ascending order of propensity score; "random", where matching takes place in a random order; "data", where matching takes place based on the order of units in the data. The matching order must be specified when using the nearest neighbor matching.
n.boot	Number of bootstrap sampling, which must be specified when method.matching="boot.optimal" or method.matching="boot.nearest". The default value is n.boot=100.
analysis.cov	Vector of names of covariates which are used for the Bayesian analysis with commensurate prior. The names of covariates must be included in lab values specified in cov.C.
method.borrow	List of information borrowing method. "noborrow" uses the concurrent data only. "fullborrow" uses the external control data without discounting. "cauchy" uses the commensurate prior to dynamically borrow the external control data, and the commensurability parameter is assumed to follow a half-Cauchy distribution. "normal" uses the commensurate prior to dynamically borrow the external control data, and the commensurability parameter is assumed to follow a half-normal distribution. "cauchy" and "normal" require to specify the scale parameter scale of half-Cauchy and half-normal distribution respectively.
chains	Number of Markov chains in MCMC sampling. The default value is chains=2.
iter	Number of iterations for each chain (including warmup) in MCMC sampling. The default value is iter=4000.
warmup	Number of warmup (burnin) iterations per chain in MCMC sampling. The default value is warmup=floor(iter/2).
thin	Period for saving samples in MCMC sampling. The default value is thin=1.
alternative	Alternative hypothesis to be tested ("greater" or "less"). The default value is alternative="greater".
sig.level	Significance level. The default value is sig.level=0.025.
nsim	Number of simulated trials.

Details

The simulation study consists of three part: data generation conducted by trial.simulation.bin function, propensity score matching conducted by psmatch function, and Bayesian analysis with

commensurate prior conducted by `commensurate.bin` function. Users can specify different sets of covariates for the propensity score matching and the Bayesian analysis.

Value

The `psborrow.bin` returns a list containing the following objects:

`reject`	Data frame containing results of Bayesian one-sided hypothesis testing (whether or not the posterior probability that the log odds ratio is greater or less than 0 exceeds 1 minus significance level): `TRUE` when significant, otherwise `FALSE`.
`theta`	Data frame containing posterior mean, median, and sd of log odds ratio.
`ov`	Data frame containing (1) overlapping coefficient of propensity score densities between treatment versus concurrent control plus external control and between concurrent control versus external control, (2) overlapping coefficient of continuous covariate densities between treatment versus concurrent control plus external control and between concurrent control versus external control, and (3) rate difference of binary covariate between treatment versus concurrent control plus external control and between concurrent control versus external control.
`n.CT`	Number of patients in treatment group in the current trial.
`n.CC`	Number of patients in concurrent control group in the current trial.
`n.ECp`	Number of patients in external control pool.
`n.EC`	Number of patients in external control.
`drift`	Odds ratio between concurrent and external control.
`true.theta`	True log odds ratio
`method.psest`	Method of estimating the propensity score.
`method.pslink`	Link function used in estimating the propensity score.
`method.whomatch`	
	Option of who to match.
`method.matching`	
	Propensity score matching method.
`method.psorder`	Order that the matching takes place when a nearest neighbor matching is used.

Examples

```
n.CT  <- 100
n.CC  <- 50
n.ECp <- 200
n.EC  <- 50

out.prob.CT <- 0.2
out.prob.CC <- 0.2
driftOR     <- 1.0

cov.C <- list(list(dist="norm",mean=0,sd=1,lab="cov1"),
              list(dist="binom",prob=0.4,lab="cov2"))

cov.cor.C <- rbind(c(  1,0.1),
```

```
                          c(0.1,  1))

cov.EC <- list(list(dist="norm",mean=0,sd=1,lab="cov1"),
               list(dist="binom",prob=0.4,lab="cov2"))

cov.cor.EC <- rbind(c(  1,0.1),
                    c(0.1,  1))

cov.effect <- c(0.9,0.9)

psmatch.cov <- c("cov1","cov2")

method.whomatch <- "conc.treat"
method.matching <- "optimal"
method.psorder  <- NULL

analysis.cov <- c("cov1")

method.borrow <- list(list(prior="noborrow"),
                      list(prior="normal",scale=0.5))

nsim <- 5

psborrow.bin(
  n.CT=n.CT, n.CC=n.CC, n.ECp=n.ECp, n.EC=n.EC,
  out.prob.CT=out.prob.CT, out.prob.CC=out.prob.CC, driftOR=driftOR,
  cov.C=cov.C, cov.cor.C=cov.cor.C,
  cov.EC=cov.EC, cov.cor.EC=cov.cor.EC, cov.effect=cov.effect,
  psmatch.cov=psmatch.cov, method.whomatch=method.whomatch,
  method.matching=method.matching, method.psorder=method.psorder,
  analysis.cov=analysis.cov, method.borrow=method.borrow,
  chains=1, iter=100, nsim=nsim)
```

---

psborrow.cont	*Simulation study of hybrid control design with Bayesian dynamic borrowing incorporating propensity score matched external control: continuous outcome*

Description

Simulation study is conducted to assess operating characteristics of hybrid control design with Bayesian dynamic borrowing, where the concurrent control is augmented by external control. The external controls are selected from external control pool using a propensity score matching. Commensurate power prior is used for Bayesian dynamic borrowing. The continuous outcome is applicable.

Usage

```
psborrow.cont(
  n.CT, n.CC, n.ECp, n.EC,
```

```
out.mean.CT, out.sd.CT, out.mean.CC, out.sd.CC, driftdiff, out.sd.EC,
cov.C, cov.cor.C, cov.EC, cov.cor.EC, cov.effect,
psmatch.cov,
method.psest="glm", method.pslink="logit",
method.whomatch, method.matching, method.psorder, n.boot=100,
analysis.cov, method.borrow,
chains=2, iter=4000, warmup=floor(iter/2), thin=1,
alternative="greater", sig.level=0.025, nsim)
```

Arguments

n.CT	Number of patients in treatment group in the current trial.
n.CC	Number of patients in concurrent control group in the current trial.
n.ECp	Number of patients in external control pool.
n.EC	Number of patients in external control.
out.mean.CT	True mean of outcome in treatment group in the current trial.
out.sd.CT	True sd of outcome in treatment group in the current trial.
out.mean.CC	True mean of outcome in concurrent control group in the current trial.
out.sd.CC	True as of outcome in concurrent control group in the current trial.
driftdiff	Mean difference between concurrent and external control for which the bias should be plotted (mean in external control minus mean in concurrent control).
out.sd.EC	True as of outcome in external control.
cov.C	List of covariate distributions for treatment and concurrent control group in the current trial. Continuous and binary covariate are applicable. The continuous covariate is assumed to follow a normal distribution; for example specified as `list(dist="norm", mean=0, sd=1, lab="cov1")`. The binary covariate is assumed to follow a binomial distribution; for example, specified as `list(dist="binom", prob=0.4, lab="cov2")`. `lab` is the column name of the covariate in the data frame generated.
cov.cor.C	Matrix of correlation coefficients for each pair of covariate for treatment and concurrent control group in the current trial, specified as Gaussian copula parameter.
cov.EC	List of covariate distributions for external control. The continuous covariate is assumed to follow a normal distribution; for example, specified as `list(dist="norm", mean=0, sd=1, lab="cov1")`. The binary covariate is assumed to follow a binomial distribution; for example, specified as `list(dist="binom", prob=0.4, lab="cov2")`. `lab` is the column name of the covariate in the data frame generated, which must be consistent with those used for `cov.C`.
cov.cor.EC	Matrix of correlation coefficients for each pair of covariate for external control, specified as Gaussian copula parameter.
cov.effect	Vector of covariate effects on the outcome, specified as mean change per one unit increase in continuous covariates or as mean change between categories for binary covariates.
psmatch.cov	Vector of names of covariates which are used for the propensity score matching. The names of covariates must be included in `lab` values specified in `cov.C`.

method.psest	Method of estimating the propensity score. Allowable options include, for example, "glm" for generalized linear model (e.g., logistic regression); "gam" for generalized additive model; "gbm" for generalized boosted model; "lasso" for lasso regression; "rpart" for classification tree. The default value is method.psest="glm".
method.pslink	Link function used in estimating the propensity score. Allowable options depend on the specific method.psest value specified. The default value is method.pslink="logit", which, along with method.psest="glm", identifies the default method as logistic regression.
method.whomatch	
	Options of who to match. Allowable options include conc.contl for matching concurrent control to external control pool; conc.treat for matching treatment to external control pool; conc.all for matching treatment plus concurrent control to external control pool; treat2contl for matching treatment to concurrent control plus external control pool.
method.matching	
	Matching method. Allowable options include "optimal" for optimal matching; "nearest" for nearest neighbor matching without replacement; "med.optimal" for equally splitting patients in the current trial and taking the median of each subset, followed by 1:1 optimal matching; "med.nearest" for equally splitting patients in the current trial and taking the median of each subset, followed by 1:1 nearest neighbor matching without replacement; "km.optimal" for k-means clustering of patients in the current trial, followed by 1:1 optimal matching; "km.nearest" for k-means clustering of patients in the current trial, followed by 1:1 nearest neighbor matching without replacement; "cm.optimal" for fuzzy c-means clustering of patients in the current trial, followed by 1:1 optimal matching; "cm.nearest" for fuzzy c-means of patients in the current trial, followed by 1:1 nearest neighbor matching without replacement; "boot.optimal" for bootstrap sampling from patients in the current trial, followed by 1:1 optimal matching; "boot.nearest" for bootstrap sampling from patient in the current trial, followed by 1:1 nearest neighbor matching without replacement.
method.psorder	Order that the matching takes place when a nearest neighbor matching is used. Allowable options include "largest", where matching takes place in descending order of propensity score; "smallest", where matching takes place in ascending order of propensity score; "random", where matching takes place in a random order; "data", where matching takes place based on the order of units in the data. The matching order must be specified when using the nearest neighbor matching.
n.boot	Number of bootstrap sampling, which must be specified when method.matching="boot.optimal" or method.matching="boot.nearest". The default value is n.boot=100.
analysis.cov	Vector of names of covariates which are used for the Bayesian analysis with commensurate prior. The names of covariates must be included in lab values specified in cov.C.
method.borrow	List of information borrowing method. "noborrow" uses the concurrent data only. "fullborrow" uses the external control data without discounting. "cauchy" uses the commensurate prior to dynamically borrow the external control data, and the commensurability parameter is assumed to follow a half-Cauchy distribution. "normal" uses the commensurate prior to dynamically borrow the

external control data, and the commensurability parameter is assumed to follow a half-normal distribution. "cauchy" and "normal" require to specify the scale parameter scale of half-Cauchy and half-normal distribution respectively.

chains	Number of Markov chains in MCMC sampling. The default value is chains=2.
iter	Number of iterations for each chain (including warmup) in MCMC sampling. The default value is iter=4000.
warmup	Number of warmup (burnin) iterations per chain in MCMC sampling. The default value is warmup=floor(iter/2).
thin	Period for saving samples in MCMC sampling. The default value is thin=1.
alternative	Alternative hypothesis to be tested ("greater" or "less"). The default value is alternative="greater".
sig.level	Significance level. The default value is sig.level=0.025.
nsim	Number of simulated trials.

Details

The simulation study consists of three part: data generation conducted by trial.simulation.cont function, propensity score matching conducted by psmatch function, and Bayesian analysis with commensurate prior conducted by commensurate.cont function. Users can specify different sets of covariates for the propensity score matching and the Bayesian analysis.

Value

The psborrow.cont returns a list containing the following objects:

reject	Data frame containing results of Bayesian one-sided hypothesis testing (whether or not the posterior probability that the mean difference is greater or less than 0 exceeds 1 minus significance level): TRUE when significant, otherwise FALSE.
theta	Data frame containing posterior mean, median, and sd of mean difference.
ov	Data frame containing (1) overlapping coefficient of propensity score densities between treatment versus concurrent control plus external control and between concurrent control versus external control, (2) overlapping coefficient of continuous covariate densities between treatment versus concurrent control plus external control and between concurrent control versus external control, and (3) rate difference of binary covariate between treatment versus concurrent control plus external control and between concurrent control versus external control.
n.CT	Number of patients in treatment group in the current trial.
n.CC	Number of patients in concurrent control group in the current trial.
n.ECp	Number of patients in external control pool.
n.EC	Number of patients in external control.
drift	Mean difference between concurrent and external control.
true.theta	True mean difference
method.psest	Method of estimating the propensity score.
method.pslink	Link function used in estimating the propensity score.

method.whomatch

> Option of who to match.

method.matching

> Propensity score matching method.

method.psorder   Order that the matching takes place when a nearest neighbor matching is used.

Examples

```
n.CT  <- 100
n.CC  <- 50
n.ECp <- 200
n.EC  <- 50

out.mean.CT <- 0
out.sd.CT   <- 1
out.mean.CC <- 0
out.sd.CC   <- 1
driftdiff   <- 0
out.sd.EC   <- 1

cov.C <- list(list(dist="norm",mean=0,sd=1,lab="cov1"),
              list(dist="binom",prob=0.4,lab="cov2"))

cov.cor.C <- rbind(c(  1,0.1),
                   c(0.1,  1))

cov.EC <- list(list(dist="norm",mean=0,sd=1,lab="cov1"),
               list(dist="binom",prob=0.4,lab="cov2"))

cov.cor.EC <- rbind(c(  1,0.1),
                    c(0.1,  1))

cov.effect <- c(0.1,0.1)

psmatch.cov <- c("cov1","cov2")

method.whomatch <- "conc.treat"
method.matching <- "optimal"
method.psorder  <- NULL

analysis.cov <- c("cov1")

method.borrow <- list(list(prior="noborrow"),
                      list(prior="normal",scale=0.5))

nsim <- 5

psborrow.cont(
  n.CT=n.CT, n.CC=n.CC, n.ECp=n.ECp, n.EC=n.EC,
  out.mean.CT=out.mean.CT, out.sd.CT=out.sd.CT,
  out.mean.CC=out.mean.CC, out.sd.CC=out.sd.CC,
  driftdiff=driftdiff, out.sd.EC=out.sd.EC,
```

```
cov.C=cov.C, cov.cor.C=cov.cor.C,
cov.EC=cov.EC, cov.cor.EC=cov.cor.EC, cov.effect=cov.effect,
psmatch.cov=psmatch.cov, method.whomatch=method.whomatch,
method.matching=method.matching, method.psorder=method.psorder,
analysis.cov=analysis.cov, method.borrow=method.borrow,
chains=1, iter=100, nsim=nsim)
```

---

psborrow.summary   *Summarizing simulation study results*

Description

Simulation study results are summarized.

Usage

```
psborrow.summary(object)
```

Arguments

object   List of simulation results.

Value

The psborrow.summary returns a list containing the following objects:

res.out   Data frame containing (1) rate that the null hypothesis is rejected in Bayesian
                one-sided hypothesis testing, (2) bias in posterior mean of treatment effect (log
                hazard ratio, log odds ratio, or mean difference), (3) empirical standard deviation
                (EmpSD), derived by taking the standard deviation of posterior mean of treatment
                effect that each simulated trial yielded, (4) model standard deviation ((ModSD)),
                derived by taking the average of posterior standard deviation of treatment ef-
                fect that each simulated trial yielded, (5) ratio of ModSD to EmpSD, (6) Coverage
                probability of credible interval.

res.cov   Data frame containing (1) mean, median, and standard deviation of overlapping
                coefficients of propensity score densities between treatment versus concurrent
                control plus external control and between concurrent control versus external
                control, (2) mean, median, and standard deviation of overlapping coefficients
                of continuous covariate densities between treatment versus concurrent control
                plus external control and between concurrent control versus external control,
                and (3) mean, median, and standard deviation of rate differences of binary co-
                variate between treatment versus concurrent control plus external control and
                between concurrent control versus external control.

n.CT   Number of patients in treatment group in the current trial.

n.CC   Number of patients in concurrent control group in the current trial.

n.ECp   Number of patients in external control pool.

n.EC   Number of patients in external control.

drift	Hazard ratio, odds ratio, or mean difference between concurrent and external control for which the bias should be plotted. The measure depends on the outcome type.
method.psest	Method of estimating the propensity score.
method.pslink	Link function used in estimating the propensity score.
method.whomatch	
	Options of who to match.
method.matching	
	Matching method.
method.psorder	Order that the matching takes place when a nearest neighbor matching is used.

---

psborrow.t2e	*Simulation study of hybrid control design with Bayesian dynamic borrowing incorporating propensity score matched external control: time-to-event outcome*

Description

Simulation study is conducted to assess operating characteristics of hybrid control design with Bayesian dynamic borrowing, where the concurrent control is augmented by external control. The external controls are selected from external control pool using a propensity score matching. Commensurate power prior is used for Bayesian dynamic borrowing. The time-to-event outcome is applicable.

Usage

```
psborrow.t2e(
  n.CT, n.CC, nevent.C, n.ECp, nevent.ECp, n.EC, accrual,
  out.mevent.CT, out.mevent.CC, driftHR,
  cov.C, cov.cor.C, cov.EC, cov.cor.EC, cov.effect,
  psmatch.cov,
  method.psest="glm", method.pslink="logit",
  method.whomatch, method.matching, method.psorder, n.boot=100,
  analysis.cov, method.borrow,
  chains=2, iter=4000, warmup=floor(iter/2), thin=1,
  alternative="greater", sig.level=0.025, nsim)
```

Arguments

n.CT	Number of patients in treatment group in the current trial.
n.CC	Number of patients in concurrent control group in the current trial.
nevent.C	Number of events in treatment and concurrent control group in the current trial.
n.ECp	Number of patients in external control pool.
nevent.ECp	Number of events in external control pool.

n.EC	Number of patients in external control.
accrual	Accrual rate, defined as the number of enrolled patients per month.
out.mevent.CT	True median time to event in treatment group in the current trial.
out.mevent.CC	True median time to event in concurrent control group in the current trial.
driftHR	Hazard ratio between concurrent and external control for which the bias should be plotted (hazard in external control divided by hazard in concurrent control).
cov.C	List of covariate distributions for treatment and concurrent control group in the current trial. Continuous and binary covariate are applicable. The continuous covariate is assumed to follow a normal distribution; for example, specified as list(dist="norm", mean=0, sd=1, lab="cov1"). The binary covariate is assumed to follow a binomial distribution; for example, specified as list(dist="binom", prob=0.4, lab="cov2"). lab is the column name of the covariate in the data frame generated.
cov.cor.C	Matrix of correlation coefficients for each pair of covariate for treatment and concurrent control group in the current trial, specified as Gaussian copula parameter.
cov.EC	List of covariate distributions for external control. The continuous covariate is assumed to follow a normal distribution; for example, specified as list(dist="norm", mean=0, sd=1, lab="cov1"). The binary covariate is assumed to follow a binomial distribution; for example, specified as list(dist="binom", prob=0.4, lab="cov2"). lab is the column name of the covariate in the data frame generated, which must be consistent with those used for cov.C.
cov.cor.EC	Matrix of correlation coefficients for each pair of covariate for external control, specified as Gaussian copula parameter.
cov.effect	Vector of covariate effects on the outcome, specified as hazard ratio per one unit increase in continuous covariate or as hazard ratio between categories for binary covariate.
psmatch.cov	Vector of names of covariates which are used for the propensity score matching. The names of covariates must be included in lab values specified in cov.C.
method.psest	Method of estimating the propensity score. Allowable options include, for example, "glm" for generalized linear model (e.g., logistic regression); "gam" for generalized additive model; "gbm" for generalized boosted model; "lasso" for lasso regression; "rpart" for classification tree. The default value is method.psest="glm".
method.pslink	Link function used in estimating the propensity score. Allowable options depend on the specific method.psest value specified. The default value is method.pslink="logit", which, along with method.psest="glm", identifies the default method as logistic regression.
method.whomatch	
	Options of who to match. Allowable options include conc.contl for matching concurrent control to external control pool; conc.treat for matching treatment to external control pool; conc.all for matching treatment plus concurrent control to external control pool; treat2contl for matching treatment to concurrent control plus external control pool.

method.matching

    Matching method. Allowable options include "optimal" for optimal matching; "nearest" for nearest neighbor matching without replacement; "med.optimal" for equally splitting patients in the current trial and taking the median of each subset, followed by 1:1 optimal matching; "med.nearest" for equally splitting patients in the current trial and taking the median of each subset, followed by 1:1 nearest neighbor matching without replacement; "km.optimal" for k-means clustering of patients in the current trial, followed by 1:1 optimal matching; "km.nearest" for k-means clustering of patients in the current trial, followed by 1:1 nearest neighbor matching without replacement; "cm.optimal" for fuzzy c-means clustering of patients in the current trial, followed by 1:1 optimal matching; "cm.nearest" for fuzzy c-means of patients in the current trial, followed by 1:1 nearest neighbor matching without replacement; "boot.optimal" for bootstrap sampling from patients in the current trial, followed by 1:1 optimal matching; "boot.nearest" for bootstrap sampling from patient in the current trial, followed by 1:1 nearest neighbor matching without replacement.

method.psorder   Order that the matching takes place when a nearest neighbor matching is used. Allowable options include "largest", where matching takes place in descending order of propensity score; "smallest", where matching takes place in ascending order of propensity score; "random", where matching takes place in a random order; "data", where matching takes place based on the order of units in the data. The matching order must be specified when using the nearest neighbor matching.

n.boot   Number of bootstrap sampling, which must be specified when method.matching="boot.optimal" or method.matching="boot.nearest". The default value is n.boot=100.

analysis.cov   Vector of names of covariates which are used for the Bayesian analysis with commensurate prior. The names of covariates must be included in lab values specified in cov.C.

method.borrow   List of information borrowing method. "noborrow" uses the concurrent data only. "fullborrow" uses the external control data without discounting. "cauchy" uses the commensurate prior to dynamically borrow the external control data, and the commensurability parameter is assumed to follow a half-Cauchy distribution. "normal" uses the commensurate prior to dynamically borrow the external control data, and the commensurability parameter is assumed to follow a half-normal distribution. "cauchy" and "normal" require to specify the scale parameter scale of half-Cauchy and half-normal distribution respectively.

chains   Number of Markov chains in MCMC sampling. The default value is chains=2.

iter   Number of iterations for each chain (including warmup) in MCMC sampling. The default value is iter=4000.

warmup   Number of warmup (burnin) iterations per chain in MCMC sampling. The default value is warmup=floor(iter/2).

thin   Period for saving samples in MCMC sampling. The default value is thin=1.

alternative   Alternative hypothesis to be tested ("greater" or "less"). The default value is alternative="greater".

sig.level   Significance level. The default value is sig.level=0.025.

nsim   Number of simulated trials.

Details

The simulation study consists of three part: data generation conducted by `trial.simulation.t2e` function, propensity score matching conducted by `psmatch` function, and Bayesian analysis with commensurate prior conducted by `commensurate.t2e` function. Users can specify different sets of covariates for the propensity score matching and the Bayesian analysis.

Value

The `psborrow.t2e` returns a list containing the following objects:

`reject`	Data frame containing results of Bayesian one-sided hypothesis testing (whether or not the posterior probability that the log hazard ratio is greater or less than 0 exceeds 1 minus significance level): `TRUE` when significant, otherwise `FALSE`.
`theta`	Data frame containing posterior mean, median, and sd of log hazard ratio.
`ov`	Data frame containing (1) overlapping coefficient of propensity score densities between treatment versus concurrent control plus external control and between concurrent control versus external control, (2) overlapping coefficient of continuous covariate densities between treatment versus concurrent control plus external control and between concurrent control versus external control, and (3) rate difference of binary covariate between treatment versus concurrent control plus external control and between concurrent control versus external control.
`n.CT`	Number of patients in treatment group in the current trial.
`n.CC`	Number of patients in concurrent control group in the current trial.
`n.ECp`	Number of patients in external control pool.
`n.EC`	Number of patients in external control.
`drift`	Hazard ratio between concurrent and external control.
`true.theta`	True log hazard ratio
`method.psest`	Method of estimating the propensity score.
`method.pslink`	Link function used in estimating the propensity score.
`method.whomatch`	
	Option of who to match.
`method.matching`	
	Propensity score matching method.
`method.psorder`	Order that the matching takes place when a nearest neighbor matching is used.

Examples

```
n.CT       <- 100
n.CC       <- 50
nevent.C   <- 100
n.ECp      <- 200
nevent.ECp <- 180
n.EC       <- 50
accrual    <- 16

out.mevent.CT <- 6
```

```
out.mevent.CC <- 6
driftHR       <- 1

cov.C <- list(list(dist="norm",mean=0,sd=1,lab="cov1"),
              list(dist="binom",prob=0.4,lab="cov2"))

cov.cor.C <- rbind(c(  1,0.1),
                   c(0.1,  1))

cov.EC <- list(list(dist="norm",mean=0,sd=1,lab="cov1"),
               list(dist="binom",prob=0.4,lab="cov2"))

cov.cor.EC <- rbind(c(  1,0.1),
                    c(0.1,  1))

cov.effect <- c(0.9,0.9)

psmatch.cov <- c("cov1","cov2")

method.whomatch <- "conc.treat"
method.matching <- "optimal"
method.psorder  <- NULL

analysis.cov <- c("cov1")

method.borrow <- list(list(prior="noborrow"),
                      list(prior="normal",scale=0.5))

nsim <- 5

psborrow.t2e(
  n.CT=n.CT, n.CC=n.CC, nevent.C=nevent.C,
  n.ECp=n.ECp, nevent.ECp=nevent.ECp, n.EC=n.EC, accrual=accrual,
  out.mevent.CT=out.mevent.CT, out.mevent.CC=out.mevent.CC, driftHR=driftHR,
  cov.C=cov.C, cov.cor.C=cov.cor.C,
  cov.EC=cov.EC, cov.cor.EC=cov.cor.EC, cov.effect=cov.effect,
  psmatch.cov=psmatch.cov, method.whomatch=method.whomatch,
  method.matching=method.matching, method.psorder=method.psorder,
  analysis.cov=analysis.cov, method.borrow=method.borrow,
  chains=1, iter=100, nsim=nsim)
```

---

psmatch   *Propensity score matching*

Description

Propensity score matching is implemented to select external controls who are more relevant to patients in the current trial with respect to covariates of interest.

Usage

```
psmatch(
  formula, data, n.EC,
  method.psest="glm", method.pslink="logit",
  method.whomatch, method.matching, method.psorder, n.boot=100)
```

Arguments

formula	Object of class formula, which is a symbolic description of the propensity score model to be fitted. The dependent variable must be named study. The explanatory variables only include covariates of interest, which must be specified in the form of linear combination.
data	Data frame, which must contain variables named study for study indicator (0 for external control, and 1 for current trial) and treat for treatment indicator (0 for concurrent and external control, and 1 for treatment).
n.EC	Number of patients in external control to be selected, which must be smaller than the number of patients in external control pool.
method.psest	Method of estimating the propensity score. Allowable options include, for example, "glm" for generalized linear model (e.g., logistic regression); "gam" for generalized additive model; "gbm" for generalized boosted model; "lasso" for lasso regression; "rpart" for classification tree. The default value is method.psest="glm".
method.pslink	Link function used in estimating the propensity score. Allowable options depend on the specific method.psest value specified. The default value is method.pslink="logit", which, along with method.psest="glm", identifies the default method as logistic regression.
method.whomatch	
	Options of who to match. Allowable options include conc.contl for matching concurrent control to external control pool; conc.treat for matching treatment to external control pool; conc.all for matching treatment plus concurrent control to external control pool; treat2contl for matching treatment to concurrent control plus external control pool.
method.matching	
	Propensity score matching method. Allowable options include "optimal" for optimal matching; "nearest" for nearest neighbor matching without replacement; "med.optimal" for equally splitting patients in the current trial and taking the median of each subset, followed by 1:1 optimal matching; "med.nearest" for equally splitting patients in the current trial and taking the median of each subset, followed by 1:1 nearest neighbor matching without replacement; "km.optimal" for k-means clustering of patients in the current trial, followed by 1:1 optimal matching; "km.nearest" for k-means clustering of patients in the current trial, followed by 1:1 nearest neighbor matching without replacement; "cm.optimal" for fuzzy c-means clustering of patients in the current trial, followed by 1:1 optimal matching; "cm.nearest" for fuzzy c-means of patients in the current trial, followed by 1:1 nearest neighbor matching without replacement; "boot.optimal" for bootstrap sampling from patients in the current trial, followed by 1:1 optimal matching; "boot.nearest" for bootstrap sampling from patient in the current trial, followed by 1:1 nearest neighbor matching without replacement.

method.psorder Order that the matching takes place when a nearest neighbor matching is used. Allowable options include `"largest"`, where matching takes place in descending order of propensity score; `"smallest"`, where matching takes place in ascending order of propensity score; `"random"`, where matching takes place in a random order; `"data"`, where matching takes place based on the order of units in the data. The matching order must be specified when using the nearest neighbor matching.

n.boot Number of bootstrap sampling, which must be specified when `method.matching="boot.optimal"` or `method.matching="boot.nearest"`. The default value is `n.boot=100`.

Details

The propensity score is defined as the conditional probability of having been included in the current trial given observed covariates. There are four options applicable for to whom the patients in external control pool are matched, including (i) concurrent control versus external control pool (`"conc.contl"`), (ii) treatment versus external control pool (`"conc.treat"`), (iii) treatment plus concurrent control versus external control pool (`"conc.all"`), and (iv) treatment versus concurrent control plus external control pool (`"treat2contl"`). Along with `method.whomatch="conc.contl"`, two 1:1 matching methods are applicable: (1) optimal matching (`"optimal"`), and (2) nearest neighbor matching without caliper (`"nearest"`). Along with `method.whomatch="conc.treat"` or `method.whomatch="conc.all"`, ten matching methods are applicable: (1) optimal matching, where 1:1 matching is first done, followed by random sampling (`"optimal"`), (2) nearest neighbor matching, where caliper is tuned iteratively to obtain the fixed number of external controls (`"nearest"`), (3) equally splitting patients in the current trial and taking the median of each subset, followed by 1:1 optimal matching (`"medm.optimal"`), (4) equally splitting patients in the current trial and taking the median of each subset, followed by 1:1 nearest neighbor matching matching (`"med.nearest"`), (5) k-means clustering of patients in the current trial, followed by 1:1 optimal matching (`"km.optimal"`), (6) k-means clustering of patients in the current trial, followed by 1:1 nearest neighbor matching (`"km.nearest"`), (7) fuzzy c-means clustering of patients in the current trial, followed by 1:1 optimal matching (`"cm.optimal"`), (8) fuzzy c-means of patients in the current trial, followed by 1:1 nearest neighbor matching (`"cm.nearest"`), (9) bootstrap sampling from patients in the current trial, followed by 1:1 optimal matching (`"boot.nearest"`), and (10) bootstrap sampling from patient in the current trial, followed by 1:1 nearest neighbor matching (`"boot.nearest"`). Along with `method.whomatch="treat2contl"`, two matching methods are applicable: (1) optimal matching, followed by random sampling (`"optimal"`), and (2) nearest neighbor matching, where caliper is tuned iteratively to obtain the fixed number of external controls (`"nearest"`).

Value

The `psmatch` returns a list containing the following objects:

subjid.EC Vector of subject ID of external control.
data.ps Data frame with estimated propensity score.

References

Austin PC. A comparison of 12 algorithms for matching on the propensity score. *Statistics in Medicine* 2014; 33:1057-1069.

Lin J, Gamalo-Siebers M, Tiwari R. Propensity score matched augmented controls in randomized clinical trials: A case study. *Pharmaceutical Statistics* 2018; 17:629-647.

Lin J, Gamalo-Siebers M, Tiwari R. Propensity-score-based priors for Bayesian augmented control design. *Pharmaceutical Statistics* 2019; 18:223-238.

Rosenbaum PR. Optimal matching for observational studies. *Journal of the American Statistical Association* 1989; 84:1024-1032.

Sawamoto R, Oba K, Matsuyama Y. Bayesian adaptive randomization design incorporating propensity score-matched historical controls. *Pharmaceutical Statistics* 2022; 21:1074-1089.

Examples

```
n.CT       <- 100
n.CC       <- 50
nevent.C   <- 100
n.ECp      <- 1000
nevent.ECp <- 800
accrual    <- 16

out.mevent.CT <- 6
out.mevent.CC <- 6
driftHR       <- 1

cov.C <- list(list(dist="norm",mean=0,sd=1,lab="cov1"),
              list(dist="binom",prob=0.4,lab="cov2"))

cov.cor.C <- rbind(c(  1,0.1),
                   c(0.1,  1))

cov.EC <- list(list(dist="norm",mean=0,sd=1,lab="cov1"),
               list(dist="binom",prob=0.4,lab="cov2"))

cov.cor.EC <- rbind(c(  1,0.1),
                    c(0.1,  1))

cov.effect <- c(0.1,0.1)

indata <- trial.simulation.t2e(
  n.CT=n.CT, n.CC=n.CC, nevent.C=nevent.C,
  n.ECp=n.ECp, nevent.ECp=nevent.ECp, accrual=accrual,
  out.mevent.CT, out.mevent.CC, driftHR,
  cov.C=cov.C, cov.cor.C=cov.cor.C,
  cov.EC=cov.EC, cov.cor.EC=cov.cor.EC, cov.effect=cov.effect)

n.EC <- 50

method.whomatch <- "conc.treat"
method.matching <- "optimal"
method.psorder  <- NULL

psmatch(
  study~cov1+cov2, data=indata, n.EC=n.EC,
```

```
      method.whomatch=method.whomatch, method.matching=method.matching,
      method.psorder=method.psorder)
```

---

trial.simulation.bin   *Simulating binary data for current trial and external control*

Description

A two-arm randomized clinical trial with a binary outcome, which is augmented by external control data, is simulated.

Usage

```
trial.simulation.bin(
 n.CT, n.CC, n.ECp,
 out.prob.CT, out.prob.CC, driftOR,
 cov.C, cov.cor.C, cov.EC, cov.cor.EC, cov.effect,
 seed=sample.int(.Machine$integer.max,1))
```

Arguments

n.CT	Number of patients in treatment group in the current trial.
n.CC	Number of patients in concurrent control group in the current trial.
n.ECp	Number of patients in external control pool.
out.prob.CT	True rate of outcome in treatment group in the current trial.
out.prob.CC	True rate of outcome in concurrent control group in the current trial.
driftOR	Odds ratio between concurrent and external control for which the bias should be plotted (odds in external control divided by odds in concurrent control).
cov.C	List of covariate distributions for treatment and concurrent control group in the current trial. Continuous and binary covariate are applicable. The continuous covariate is assumed to follow a normal distribution; for example, specified as list(dist="norm", mean=0, sd=1, lab="cov1"). The binary covariate is assumed to follow a binomial distribution; for example, specified as list(dist="binom", prob=0.4, lab="cov2"). lab is the column name of the covariate in the data frame generated.
cov.cor.C	Matrix of correlation coefficients for each pair of covariate for treatment and concurrent control in the current trial, specified as Gaussian copula parameter.
cov.EC	List of covariate distributions for external control. The continuous covariate is assumed to follow a normal distribution; for example, specified as list(dist="norm", mean=0, sd=1, lab="cov1"). The binary covariate is assumed to follow a binomial distribution; for example, specified as list(dist="binom", prob=0.4, lab="cov2"). lab is the column name of the covariate in the data frame generated, which must be consistent with those used for cov.C.
cov.cor.EC	Matrix of correlation coefficients for each pair of covariate for external control, specified as Gaussian copula parameter.

cov.effect   Vector of covariate effects on the outcome, specified as odds ratio per one unit
                  increase in continuous covariates or as odds ratio between categories for binary
                  covariates.

seed   Setting a seed.

Details

The binary outcome is assumed to follow a binomial distribution. Given more than one covariates
with their effects on the outcome, a logistic regression model is constructed for data generation. The
data frame generated include the binary outcome data and covariates for `n.CT` and `n.CC` patients
in treatment and concurrent control group in the current trial respectively, and `n.ECp` patients in
external control pool. One record per patient. More than one covariates must be specified.

Value

The `trial.simulation.bin` returns a data frame containing the following variables:

study   Study indicator (0 for external control, and 1 for current trial)

treat   Treatment indicator (0 for concurrent and external control, and 1 for treatment)

y   Binary outcome

column name `specified`
                  Covariate of interest

Examples

```
n.CT  <- 100
n.CC  <- 50
n.ECp <- 1000

out.prob.CT <- 0.2
out.prob.CC <- 0.2
driftOR     <- 1.0

cov.C <- list(list(dist="norm",mean=0,sd=1,lab="cov1"),
              list(dist="binom",prob=0.4,lab="cov2"))

cov.cor.C <- rbind(c(  1,0.1),
                   c(0.1,  1))

cov.EC <- list(list(dist="norm",mean=0,sd=1,lab="cov1"),
               list(dist="binom",prob=0.4,lab="cov2"))

cov.cor.EC <- rbind(c(  1,0.1),
                    c(0.1,  1))

cov.effect <- c(0.8,0.8)

trial.simulation.bin(
   n.CT=n.CT, n.CC=n.CC, n.ECp=n.ECp,
   out.prob.CT=out.prob.CT, out.prob.CC=out.prob.CC, driftOR=driftOR,
   cov.C=cov.C, cov.cor.C=cov.cor.C,
```

```
    cov.EC=cov.EC, cov.cor.EC=cov.cor.EC, cov.effect=cov.effect, seed=100)
```

---

trial.simulation.cont *Simulating continuous data for current trial and external control*

Description

A two-arm randomized clinical trial with a continuous outcome, which is augmented by external control data, is simulated.

Usage

```
trial.simulation.cont(
 n.CT, n.CC, n.ECp,
 out.mean.CT, out.sd.CT, out.mean.CC, out.sd.CC, driftdiff, out.sd.EC,
 cov.C, cov.cor.C, cov.EC, cov.cor.EC, cov.effect,
 seed=sample.int(.Machine$integer.max,1))
```

Arguments

n.CT	Number of patients in treatment group in the current trial.
n.CC	Number of patients in concurrent control group in the current trial.
n.ECp	Number of patients in external control pool.
out.mean.CT	True mean of outcome in treatment group in the current trial.
out.sd.CT	True sd of outcome in treatment group in the current trial.
out.mean.CC	True mean of outcome in concurrent control group in the current trial.
out.sd.CC	True sd of outcome in concurrent control group in the current trial.
driftdiff	Mean difference between concurrent and external control for which the bias should be plotted (mean in external control minus mean in concurrent control).
out.sd.EC	True sd of outcome in external control.
cov.C	List of covariate distributions for treatment and concurrent control group in the current trial. Continuous and binary covariate are applicable. The continuous covariate is assumed to follow a normal distribution; for example specified as list(dist="norm", mean=0, sd=1, lab="cov1"). The binary covariate is assumed to follow a binomial distribution; for example, specified as list(dist="binom", prob=0.4, lab="cov2"). lab is the column name of the covariate in the data frame generated.
cov.cor.C	Matrix of correlation coefficients for each pair of covariate for treatment and concurrent control group in the current trial, specified as Gaussian copula parameter.
cov.EC	List of covariate distributions for external control. The continuous covariate is assumed to follow a normal distribution; for example, specified as list(dist="norm", mean=0, sd=1, lab="cov1"). The binary covariate is assumed to follow a binomial distribution; for example, specified as list(dist="binom", prob=0.4, lab="cov2"). lab is the column name of the covariate in the data frame generated, which must be consistent with those used for cov.C.

cov.cor.EC	Matrix of correlation coefficients for each pair of covariate for external control, specified as Gaussian copula parameter.
cov.effect	Vector of covariate effects on the outcome, specified as mean change per one unit increase in continuous covariates or as mean change between categories for binary covariates.
seed	Setting a seed.

Details

The continuous outcome is assumed to follow a normal distribution. Given more than one covariates with their effects on the outcome, a normal linear regression model is constructed for data generation. The data frame generated include the continuous outcome data and covariates for n.CT and n.CC patients in treatment and concurrent control group in the current trial respectively, and n.ECp patients in external control pool. One record per patient. More than one covariates must be specified.

Value

The trial.simulation.cont returns a data frame containing the following variables:

study	Study indicator (0 for external control, and 1 for current trial)
treat	Treatment indicator (0 for concurrent and external control, and 1 for treatment)
y	Continuous outcome
column name specified	
	Covariate of interest

Examples

```
n.CT  <- 100
n.CC  <- 50
n.ECp <- 1000

out.mean.CT <- 0
out.sd.CT   <- 1
out.mean.CC <- 0
out.sd.CC   <- 1
driftdiff   <- 0
out.sd.EC   <- 1

cov.C <- list(list(dist="norm",mean=0,sd=1,lab="cov1"),
              list(dist="binom",prob=0.4,lab="cov2"))

cov.cor.C <- rbind(c(  1,0.1),
                   c(0.1,  1))

cov.EC <- list(list(dist="norm",mean=0,sd=1,lab="cov1"),
               list(dist="binom",prob=0.4,lab="cov2"))

cov.cor.EC <- rbind(c(  1,0.1),
                    c(0.1,  1))
```

```
cov.effect <- c(0.1,0.1)

trial.simulation.cont(
  n.CT=n.CT, n.CC=n.CC, n.ECp=n.ECp,
  out.mean.CT=out.mean.CT, out.sd.CT=out.sd.CT,
  out.mean.CC=out.mean.CC, out.sd.CC=out.sd.CC,
  driftdiff=driftdiff, out.sd.EC=out.sd.EC,
  cov.C=cov.C, cov.cor.C=cov.cor.C,
  cov.EC=cov.EC, cov.cor.EC=cov.cor.EC, cov.effect=cov.effect, seed=100)
```

---

trial.simulation.t2e *Simulating time-to-event data for current trial and external control*

Description

A two-arm randomized clinical trial with a time-to-event outcome, which is augmented by external control data, is simulated.

Usage

```
trial.simulation.t2e(
 n.CT, n.CC, nevent.C, n.ECp, nevent.ECp, accrual,
 out.mevent.CT, out.mevent.CC, driftHR,
 cov.C, cov.cor.C, cov.EC, cov.cor.EC, cov.effect,
 seed=sample.int(.Machine$integer.max,1))
```

Arguments

n.CT	Number of patients in treatment group in the current trial.
n.CC	Number of patients in concurrent control group in the current trial.
nevent.C	Number of events in treatment and concurrent control group in the current trial.
n.ECp	Number of patients in external control pool.
nevent.ECp	Number of events in external control pool.
accrual	Accrual rate, defined as the number of enrolled patients per month.
out.mevent.CT	True median time to event in treatment group in the current trial.
out.mevent.CC	True median time to event in concurrent control group in the current trial.
driftHR	Hazard ratio between concurrent and external control for which the bias should be plotted (hazard in external control divided by hazard in concurrent control).
cov.C	List of covariate distributions for treatment and concurrent control group in the current trial. Continuous and binary covariate are applicable. The continuous covariate is assumed to follow a normal distribution; for example, specified as list(dist="norm", mean=0, sd=1, lab="cov1"). The binary covariate is assumed to follow a binomial distribution; for example, specified as list(dist="binom", prob=0.4, lab="cov2"). lab is the column name of the covariate in the data frame generated.

cov.cor.C	Matrix of correlation coefficients for each pair of covariate for treatment and concurrent control group in the current trial, specified as Gaussian copula parameter.
cov.EC	List of covariate distributions for external control. The continuous covariate is assumed to follow a normal distribution; for example, specified as `list(dist="norm", mean=0, sd=1, lab="cov1")`. The binary covariate is assumed to follow a binomial distribution; for example, specified as `list(dist="binom", prob=0.4, lab="cov2")`. `lab` is the column name of the covariate in the data frame generated, which must be consistent with those used for `cov.C`.
cov.cor.EC	Matrix of correlation coefficients for each pair of covariate for external control, specified as Gaussian copula parameter.
cov.effect	Vector of covariate effects on the outcome , specified as hazard ratio per one unit increase in continuous covariates or as hazard ratio between categories for binary covariates.
seed	Setting a seed.

Details

The time to event outcome is assumed to follow a Weibull distribution. Given more than one covariates with their effects on the outcome, a Weibull proportional hazards model is constructed for data generation. The data frame generated include the time-to-event outcome data and covariates for `n.CT` and `n.CC` patients in treatment and concurrent control group in the current trial respectively, and `n.ECp` patients in external control pool. One record per patient. More than one covariates must be specified.

Value

The `trial.simulation.t2e` returns a data frame containing the following variables:

study	Study indicator (0 for external control, and 1 for current trial)
treat	Treatment indicator (0 for concurrent and external control, and 1 for treatment)
time	Time to event or censoring
status	Censoring (0 for censored, and 1 for event occurred)
column name `specified`	
	Covariate of interest

Examples

```
n.CT       <- 100
n.CC       <- 50
nevent.C   <- 100
n.ECp      <- 1000
nevent.ECp <- 800
accrual    <- 16

out.mevent.CT <- 6
out.mevent.CC <- 6
driftHR       <- 1
```

```
cov.C <- list(list(dist="norm",mean=0,sd=1,lab="cov1"),
              list(dist="binom",prob=0.4,lab="cov2"))

cov.cor.C <- rbind(c(  1,0.1),
                   c(0.1,  1))

cov.EC <- list(list(dist="norm",mean=0,sd=1,lab="cov1"),
               list(dist="binom",prob=0.4,lab="cov2"))

cov.cor.EC <- rbind(c(  1,0.1),
                    c(0.1,  1))

cov.effect <- c(0.8,0.8)

trial.simulation.t2e(
   n.CT=n.CT, n.CC=n.CC, nevent.C=nevent.C,
   n.ECp=n.ECp, nevent.ECp=nevent.ECp, accrual=accrual,
   out.mevent.CT, out.mevent.CC, driftHR,
   cov.C=cov.C, cov.cor.C=cov.cor.C,
   cov.EC=cov.EC, cov.cor.EC=cov.cor.EC, cov.effect=cov.effect, seed=100)
```

Index