

Package ‘trtswitch’

October 17, 2024

Type Package

Title Treatment Switching

Version 0.1.0

Date 2024-10-15

Description Implements rank-preserving structural failure time model (RPSFTM), iterative parameter estimation (IPE), inverse probability of censoring weights (IPCW), and two-stage estimation (TSE) methods for treatment switching in randomized clinical trials.

License GPL (>= 2)

Imports Rcpp (>= 1.0.9)

LinkingTo Rcpp

Suggests testthat (>= 3.0.0), dplyr (>= 1.1.4), survival (>= 3.5-7),
geepack (>= 1.3.12)

RoxygenNote 7.3.1

Encoding UTF-8

NeedsCompilation yes

Depends R (>= 2.10)

LazyData true

Config/testthat/edition 3

Author Kaifeng Lu [aut, cre] (<<https://orcid.org/0000-0002-6160-7119>>)

Maintainer Kaifeng Lu <kaifenglu@gmail.com>

Repository CRAN

Date/Publication 2024-10-17 10:10:03 UTC

Contents

trtswitch-package	2
aml	3
bscpp	4
heart	5
immdef	5

ingots	6
ipcw	7
ipe	11
kmest	15
liferegr	17
logisregr	21
lrtest	24
nscpp	26
phregr	27
qrcpp	31
rawdata	32
residuals_phregr	33
rmdiff	34
rmest	36
rpsftm	37
sexagg	41
shilong	42
six	43
splineDesigncpp	44
survfit_phregr	45
tobin	47
tsegest	47
tsegestsim	52
tsesimp	56

Index	60
--------------	-----------

trtswitch-package	<i>Treatment Switching</i>
-------------------	----------------------------

Description

Implements rank-preserving structural failure time model (RPSFTM), iterative parameter estimation (IPE), inverse probability of censoring weights (IPCW), and two-stage estimation (TSE) methods for treatment switching in randomized clinical trials.

Details

To enable bootstrapping of the parameter estimates, we implements many standard survival analysis methods in C++. These include but are not limited to Kaplan-Meier estimates of the survival curves, log-rank tests, accelerate failure time models, and Cox proportional hazards models.

All three treatment switching adjustment methods allow treatment switching in both treatment arms, stratification and covariates adjustment. For the AFT models, stratification factors are included as covariates (main effects only or all-way interactions) because SAS PROC LIFEREG does not have the strata statement. The RPSFTM, IPE and TSE methods can be used with or without recensoring. The IPCW method can produce stabilized and truncated weights.

Currently, for the IPCW method, only the Cox model with time-dependent covariates is used to fit the switching model. For the TSE method, only the simple method assuming no gap between disease progression and treatment switching is implemented.

Author(s)

Kaifeng Lu, <kaifenglu@gmail.com>

References

- James M. Robins and Anastasios A. Tsiatis. Correcting for non-compliance in randomized trials using rank preserving structural failure time models. *Communications in Statistics*. 1991;20(8):2609-2631.
- Ian R. White, Abdel G. Babiker, Sarah Walker, and Janet H. Darbyshire. Randomization-based methods for correcting for treatment changes: Examples from the CONCORDE trial. *Statistics in Medicine*. 1999;18:2617-2634.
- Michael Branson and John Whitehead. Estimating a treatment effect in survival studies in which patients switch treatment. *Statistics in Medicine*. 2002;21:2449-2463.
- Ian R White. Letter to the Editor: Estimating treatment effects in randomized trials with treatment switching. *Statistics in Medicine*. 2006;25:1619-1622.
- James M. Robins and Dianne M. Finkelstein. Correcting for noncompliance and dependent censoring in an AIDS clinical trial with inverse probability of censoring weighted (IPCW) log-rank tests. *Biometrics*. 2000;56:779-788.
- Nicholas R Latimer, KR Abrams, PC Lambert, MK Crowther, AJ Wailoo, JP Morden, RL Akehurst, and MJ Campbell. Adjusting for treatment switching in randomised controlled trials - A simulation study and a simplified two-stage method. *Statistical Methods in Medical Research*. 2017;26(2):724-751.

aml

Acute myelogenous leukemia survival data from the survival package

Description

Survival in patients with acute myelogenous leukemia.

time Survival or censoring time

status censoring status

x maintenance chemotherapy given or not

Usage

aml

Format

An object of class `data.frame` with 23 rows and 3 columns.

Description

Computes the B-spline basis matrix for a given polynomial spline.

Usage

```
bscpp(
  x = NA_real_,
  df = NA_integer_,
  knots = NA_real_,
  degree = 3L,
  intercept = 0L,
  boundary_knots = NA_real_,
  warn_outside = 1L
)
```

Arguments

<code>x</code>	A numeric vector representing the predictor variable.
<code>df</code>	Degrees of freedom, specifying the number of columns in the basis matrix. If <code>df</code> is provided, the function automatically selects <code>df - degree - intercept</code> internal knots based on appropriate quantiles of <code>x</code> , ignoring any missing values.
<code>knots</code>	A numeric vector specifying the internal breakpoints that define the spline. If not provided, <code>df</code> must be specified.
<code>degree</code>	An integer specifying the degree of the piecewise polynomial. The default value is 3, which corresponds to cubic splines.
<code>intercept</code>	A logical value indicating whether to include an intercept in the basis. The default is <code>FALSE</code> .
<code>boundary_knots</code>	A numeric vector of length 2 specifying the boundary points where the B-spline basis should be anchored. If not supplied, the default is the range of non-missing values in <code>x</code> .
<code>warn_outside</code>	A logical value indicating whether a warning should be issued if any values of <code>x</code> fall outside the specified boundary knots.

Value

A matrix with dimensions `c(length(x), df)`. If `df` is provided, the matrix will have `df` columns. Alternatively, if `knots` are supplied, the number of columns will be `length(knots) + degree + intercept`. The matrix contains attributes that correspond to the arguments passed to the `bscpp` function.

Author(s)

Kaifeng Lu, <kaifenglu@gmail.com>

Examples

```
bscpp(women$height, df = 5)
```

heart

Stanford heart transplant data from the survival package

Description

Survival of patients on the waiting list for the Stanford heart transplant program.

start, stop, event entry and exit time and status for the time interval

age age-48 years

year year of acceptance (in years after Nov 1, 1967)

surgery prior bypass surgery 1=yes, 0=no

transplant received transplant 1=yes, 0=no

id patient id

Usage

```
heart
```

Format

An object of class `data.frame` with 172 rows and 8 columns.

immdef

Simulated CONCORDE trial data from the rpsftm package

Description

Patients were randomly assigned to receive treatment immediately or deferred, and those in the deferred arm could cross over and receive treatment. The primary endpoint was time to disease progression.

id Patient identification number

def Indicator that the participant was assigned to the deferred treatment arm

imm Indicator that the participant was assigned to the immediate treatment arm

censyrs The censoring time, in years, corresponding to the close of study minus the time of entry for each patient

xo Indicator that crossover occurred
 xoyrs The time, in years, from entry to switching, or 0 for patients in the immediate arm
 prog Indicator of disease progression (1), or censoring (0)
 progyrs Time, in years, from entry to disease progression or censoring
 entry The time of entry into the study, measured in years from the date of randomisation

Usage

immdef

Format

An object of class `data.frame` with 1000 rows and 9 columns.

ingots *The binary data from Cox and Snell (1989, pp. 10-11).*

Description

The dataset consists of the number of ingots not ready for rolling and the number of ingots ready for rolling for a number of combinations of heating time and soaking time.

Usage

ingots

Format

An object of class `tbl_df` (inherits from `tbl`, `data.frame`) with 25 rows and 4 columns.

Details

Heat The heating time

Soak The soaking time

NotReady Response indicator, with a value 1 for units not ready for rolling (event) and a value of 0 for units ready for rolling (nonevent)

Freq The frequency of occurrence of each combination of Heat, Soak, and NotReady

ipcw	<i>Inverse Probability of Censoring Weights (IPCW) Method for Treatment Switching</i>
------	---

Description

Uses the IPCW method to obtain the hazard ratio estimate of the Cox model to account for treatment switching.

Usage

```
ipcw(  
  data,  
  id = "id",  
  stratum = "",  
  tstart = "tstart",  
  tstop = "tstop",  
  event = "event",  
  treat = "treat",  
  swtrt = "swtrt",  
  swtrt_time = "swtrt_time",  
  swtrt_time_lower = "",  
  swtrt_time_upper = "",  
  base_cov = "",  
  numerator = "",  
  denominator = "",  
  logistic_switching_model = FALSE,  
  strata_main_effect_only = TRUE,  
  firth = FALSE,  
  flic = FALSE,  
  ns_df = 3,  
  stabilized_weights = TRUE,  
  trunc = 0,  
  trunc_upper_only = TRUE,  
  swtrt_control_only = TRUE,  
  alpha = 0.05,  
  ties = "efron",  
  boot = TRUE,  
  n_boot = 1000,  
  seed = NA  
)
```

Arguments

data	The input data frame that contains the following variables: <ul style="list-style-type: none">• id: The id to identify observations belonging to the same subject for counting process data with time-dependent covariates.
------	---

- `stratum`: The stratum.
- `tstart`: The starting time of the time interval for counting-process data with time-dependent covariates.
- `tstop`: The stopping time of the time interval for counting-process data with time-dependent covariates.
- `event`: The event indicator, 1=event, 0=no event.
- `treat`: The randomized treatment indicator, 1=treatment, 0=control.
- `swtrt`: The treatment switch indicator, 1=switch, 0=no switch.
- `swtrt_time`: The time from randomization to treatment switch.
- `swtrt_time_lower`: The lower bound of treatment switching time.
- `swtrt_time_upper`: The upper bound of treatment switching time.
- `base_cov`: The baseline covariates (excluding `treat`) used in the outcome model.
- `numerator`: The baseline covariates used in the switching model for the numerator for stabilized weights.
- `denominator`: The baseline and time-dependent covariates used in the switching model for the denominator.

<code>id</code>	The name of the <code>id</code> variable in the input data.
<code>stratum</code>	The name(s) of the <code>stratum</code> variable(s) in the input data.
<code>tstart</code>	The name of the <code>tstart</code> variable in the input data.
<code>tstop</code>	The name of the <code>tstop</code> variable in the input data.
<code>event</code>	The name of the <code>event</code> variable in the input data.
<code>treat</code>	The name of the <code>treatment</code> variable in the input data.
<code>swtrt</code>	The name of the <code>swtrt</code> variable in the input data.
<code>swtrt_time</code>	The name of the <code>swtrt_time</code> variable in the input data.
<code>swtrt_time_lower</code>	The name of the <code>swtrt_time_lower</code> variable in the input data.
<code>swtrt_time_upper</code>	The name of the <code>swtrt_time_upper</code> variable in the input data.
<code>base_cov</code>	The vector of names of <code>base_cov</code> variables (excluding <code>treat</code>) in the input data for the Cox model.
<code>numerator</code>	The vector of names of variables in the input data for the numerator switching model for stabilized weights.
<code>denominator</code>	The vector of names of variables in the input data for the denominator switching model.
<code>logistic_switching_model</code>	Whether a pooled logistic regression switching model is used.
<code>strata_main_effect_only</code>	Whether to only include the strata main effects in the logistic regression switching model. Defaults to TRUE, otherwise all possible strata combinations will be considered in the switching model.
<code>firth</code>	Whether the firth's bias reducing penalized likelihood should be used. The default is FALSE.

flic	Whether to apply intercept correction to obtain more accurate predicted probabilities. The default is FALSE.
ns_df	Degrees of freedom for the natural cubic spline. Defaults to 3 for two inner knots at the 33 and 67 percentiles of the artificial censoring times due to treatment switching.
stabilized_weights	Whether to use the stabilized weights.
trunc	The pre-specified fraction of the weights. Defaults to 0 for no truncation in weights.
trunc_upper_only	Whether to truncate the weights from the upper end of the distribution only. Defaults to TRUE, otherwise the weights will be truncated from both the lower and upper ends of the distribution.
swrt_control_only	Whether treatment switching occurred only in the control group.
alpha	The significance level to calculate confidence intervals.
ties	The method for handling ties in the Cox model, either "breslow" or "efron" (default).
boot	Whether to use bootstrap to obtain the confidence interval for hazard ratio. Defaults to TRUE.
n_boot	The number of bootstrap samples.
seed	The seed to reproduce the bootstrap results. The seed from the environment will be used if left unspecified.

Details

We use the following steps to obtain the hazard ratio estimate and confidence interval had there been no treatment switching:

- Exclude observations after treatment switch and set up the crossover and event indicators for the last time interval for each subject.
- For time-dependent covariates Cox switching models, replicate unique event times across treatment arms within each subject.
- Fit the denominator switching model (and the numerator switching model for stabilized weights) to obtain the inverse probability of censoring weights. This can be a Cox model with time-dependent covariates or a pooled logistic regression model.
- Fit the weighted Cox model to the censored outcome survival times to obtain the hazard ratio estimate.
- Use bootstrap to construct the p-value and confidence interval for hazard ratio.

Value

A list with the following components:

- logrank_pvalue: The two-sided p-value of the log-rank test based on the treatment policy strategy.

- `cox_pvalue`: The two-sided p-value for treatment effect based on the Cox model.
- `hr`: The estimated hazard ratio from the Cox model.
- `hr_CI`: The confidence interval for hazard ratio.
- `hr_CI_type`: The type of confidence interval for hazard ratio, either "Cox model" or "bootstrap".
- `fit_switch`: A list of the fitted switching models for the denominator and numerator by treatment group.
- `df_outcome`: The input data frame for the outcome Cox model including the inverse probability of censoring weights.
- `fit_outcome`: The fitted outcome model.
- `settings`: A list with the following components:
 - `logistic_switching_model`: Whether a pooled logistic regression switching model is used.
 - `strata_main_effect_only`: Whether to only include the strata main effects in the logistic regression switching model.
 - `firth`: Whether the firth's bias reducing penalized likelihood should be used.
 - `flic`: Whether to apply intercept correction to obtain more accurate predicted probabilities.
 - `stabilized_weights`: Whether to use the stabilized weights.
 - `trunc`: The truncation fraction of the weight distribution.
 - `trunc_upper_only`: Whether to truncate the weights from the upper end of the distribution only.
 - `swtrt_control_only`: Whether treatment switching occurred only in the control group.
 - `alpha`: The significance level to calculate confidence intervals.
 - `ties`: The method for handling ties in the Cox model.
 - `boot`: Whether to use bootstrap to obtain the confidence interval for hazard ratio.
 - `n_boot`: The number of bootstrap samples.
 - `seed`: The seed to reproduce the simulation results.
- `hr_boots`: The bootstrap hazard ratio estimates if `boot` is TRUE.

Author(s)

Kaifeng Lu, <kaifenglu@gmail.com>

References

James M. Robins and Dianne M. Finkelstein. Correcting for noncompliance and dependent censoring in an AIDS clinical trial with inverse probability of censoring weighted (IPCW) log-rank tests. *Biometrics*. 2000;56:779-788.

Examples

```

# Example 1: pooled logistic regression switching model

library(dplyr)

sim1 <- tsegestsim(
  n = 500, allocation1 = 2, allocation2 = 1, pbprog = 0.5,
  trtlghr = -0.5, bprogs1 = 0.3, shape1 = 1.8,
  scale1 = 0.000025, shape2 = 1.7, scale2 = 0.000015,
  pmix = 0.5, admin = 5000, pcatnotrtbprog = 0.5,
  pcatrtbprog = 0.25, pcatnotrt = 0.2, pcatrt = 0.1,
  catmult = 0.5, tdxo = 1, ppoor = 0.1, pgood = 0.04,
  ppoormet = 0.4, pgoodmet = 0.2, xomult = 1.4188308,
  milestone = 546, swtrt_control_only = TRUE,
  outputRawDataset = 1, seed = 2000)

fit1 <- ipcw(
  sim1$paneldata, id = "id", tstart = "tstart",
  tstop = "tstop", event = "died", treat = "trtrand",
  swtrt = "xo", swtrt_time = "xotime",
  swtrt_time_lower = "timePFSobs",
  swtrt_time_upper = "xotime_upper", base_cov = "bprog",
  numerator = "bprog", denominator = "bprog*catlag",
  logistic_switching_model = TRUE, ns_df = 3,
  swtrt_control_only = TRUE, boot = FALSE)

c(fit1$hr, fit1$hr_CI)

# Example 2: time-dependent covariates Cox switching model

fit2 <- ipcw(
  shilong, id = "id", tstart = "tstart", tstop = "tstop",
  event = "event", treat = "bras.f", swtrt = "co",
  swtrt_time = "dco",
  base_cov = c("agerand", "sex.f", "tt_Lnum", "rmh_alea.c",
              "pathway.f"),
  numerator = c("agerand", "sex.f", "tt_Lnum", "rmh_alea.c",
              "pathway.f"),
  denominator = c("agerand", "sex.f", "tt_Lnum", "rmh_alea.c",
              "pathway.f", "ps", "ttc", "tran"),
  swtrt_control_only = FALSE, boot = FALSE)

c(fit2$hr, fit2$hr_CI)

```

Description

Obtains the causal parameter estimate from the accelerated failure-time (AFT) model and the hazard ratio estimate from the Cox model to account for treatment switching.

Usage

```
ipe(
  data,
  stratum = "",
  time = "time",
  event = "event",
  treat = "treat",
  rx = "rx",
  censor_time = "censor_time",
  base_cov = "",
  aft_dist = "weibull",
  strata_main_effect_only = 1,
  treat_modifier = 1,
  recensor = TRUE,
  admin_recensor_only = FALSE,
  autoswitch = TRUE,
  alpha = 0.05,
  ties = "efron",
  tol = 1e-06,
  boot = FALSE,
  n_boot = 1000,
  seed = NA
)
```

Arguments

<code>data</code>	The input data frame that contains the following variables: <ul style="list-style-type: none"> • <code>stratum</code>: The stratum. • <code>time</code>: The survival time for right censored data. • <code>event</code>: The event indicator, 1=event, 0=no event. • <code>treat</code>: The randomized treatment indicator, 1=treatment, 0=control. • <code>rx</code>: The proportion of time on active treatment. • <code>censor_time</code>: The administrative censoring time. It should be provided for all subjects including those who had events. • <code>base_cov</code>: The baseline covariates (excluding treat).
<code>stratum</code>	The name(s) of the stratum variable(s) in the input data.
<code>time</code>	The name of the time variable in the input data.
<code>event</code>	The name of the event variable in the input data.
<code>treat</code>	The name of the treatment variable in the input data.
<code>rx</code>	The name of the rx variable in the input data.

censor_time	The name of the censor_time variable in the input data.
base_cov	The vector of names of baseline covariates (excluding treat) in the input data.
aft_dist	The assumed distribution for time to event for the AFT model. Options include "exponential", "weibull", "loglogistic", and "lognormal".
strata_main_effect_only	Whether to only include the strata main effects in the AFT model. Defaults to TRUE, otherwise all possible strata combinations will be considered in the AFT model.
treat_modifier	The optional sensitivity parameter for the constant treatment effect assumption.
recensor	Whether to apply recensoring to counter-factual survival times. Defaults to TRUE.
admin_recensor_only	Whether to apply recensoring to administrative censoring time only. Defaults to FALSE, in which case, recensoring will be applied to the actual censoring time for dropouts.
autoswitch	Whether to exclude recensoring for treatment arms with no switching. Defaults to TRUE.
alpha	The significance level to calculate confidence intervals.
ties	The method for handling ties in the Cox model, either "breslow" or "efron" (default).
tol	The desired accuracy (convergence tolerance) for psi.
boot	Whether to use bootstrap to obtain the confidence interval for hazard ratio. Defaults to FALSE, in which case, the confidence interval will be constructed to match the log-rank test p-value.
n_boot	The number of bootstrap samples.
seed	The seed to reproduce the simulation results. The seed from the environment will be used if left unspecified.

Details

We use the following steps to obtain the hazard ratio estimate and confidence interval had there been no treatment switching:

- Use IPE to estimate the causal parameter ψ based on the AFT model for counter-factual survival times for both arms: $U_{i,\psi} = T_{C_i} + e^{\psi} T_{E_i}$ for the control arm, and $V_{i,\psi} = T_{E_i} + e^{-\psi} T_{C_i}$ for the experimental arm.
- Fit the Cox proportional hazards model to the counter-factual survival times to obtain the hazard ratio estimate.
- Use either the log-rank test p-value for the treatment policy strategy or bootstrap to construct the confidence interval for hazard ratio.

Value

A list with the following components:

- psi: The estimated causal parameter.

- `psi_CI`: The confidence interval for `psi`.
- `psi_CI_type`: The type of confidence interval for `psi`, i.e., "log-rank p-value" or "bootstrap".
- `Sstar`: A data frame containing the counter-factual unswitched survival times and the event indicators.
- `kmstar`: A data frame containing the Kaplan-Meier estimates based on the counter-factual unswitched survival times by treatment arm.
- `logrank_pvalue`: The two-sided p-value of the log-rank test based on the treatment policy strategy.
- `cox_pvalue`: The two-sided p-value for treatment effect based on the Cox model.
- `hr`: The estimated hazard ratio from the Cox model.
- `hr_CI`: The confidence interval for hazard ratio.
- `hr_CI_type`: The type of confidence interval for hazard ratio, either "log-rank p-value" or "bootstrap".
- `settings`: A list with the following components:
 - `aft_dist`: The distribution for time to event for the AFT model.
 - `strata_main_effect_only`: Whether to only include the strata main effects in the AFT model.
 - `treat_modifier`: The sensitivity parameter for the constant treatment effect assumption.
 - `recensor`: Whether to apply recensoring to counter-factual survival times.
 - `admin_recensor_only`: Whether to apply recensoring to administrative censoring time only.
 - `autoswitch`: Whether to exclude recensoring for treatment arms with no switching.
 - `alpha`: The significance level to calculate confidence intervals.
 - `ties`: The method for handling ties in the Cox model.
 - `tol`: The desired accuracy (convergence tolerance).
 - `boot`: Whether to use bootstrap to obtain the confidence interval for hazard ratio.
 - `n_boot`: The number of bootstrap samples.
 - `seed`: The seed to reproduce the bootstrap results.
- `hr_boots`: The bootstrap hazard ratio estimates if `boot` is TRUE.
- `psi_boots`: The bootstrap `psi` estimates if `boot` is TRUE.

Author(s)

Kaifeng Lu, <kaifenglu@gmail.com>

References

Michael Branson and John Whitehead. Estimating a treatment effect in survival studies in which patients switch treatment. *Statistics in Medicine*. 2002;21:2449-2463.

Ian R White. Letter to the Editor: Estimating treatment effects in randomized trials with treatment switching. *Statistics in Medicine*. 2006;25:1619-1622.

Examples

```

library(dplyr)

# Example 1: one-way treatment switching (control to active)

data <- immdef %>% mutate(rx = 1-xoyrs/progyrs)

fit1 <- ipe(
  data, time = "progyrs", event = "prog", treat = "imm",
  rx = "rx", censor_time = "censyrs", aft_dist = "weibull",
  boot = FALSE)

c(fit1$hr, fit1$hr_CI)

# Example 2: two-way treatment switching (illustration only)

# the eventual survival time
shilong1 <- shilong %>%
  arrange(bras.f, id, tstop) %>%
  group_by(bras.f, id) %>%
  filter(row_number() == n()) %>%
  select(-c("ps", "ttc", "tran"))

shilong2 <- shilong1 %>%
  mutate(rx = ifelse(co, ifelse(bras.f == "MTA", dco/ady,
                               1 - dco/ady),
                    ifelse(bras.f == "MTA", 1, 0)))

fit2 <- ipe(
  shilong2, time = "tstop", event = "event",
  treat = "bras.f", rx = "rx", censor_time = "dcut",
  base_cov = c("agerand", "sex.f", "tt_Lnum", "rmh_alea.c",
              "pathway.f"),
  aft_dist = "weibull", boot = FALSE)

c(fit2$hr, fit2$hr_CI)

```

kmest

Kaplan-Meier Estimates of Survival Curve

Description

Obtains the Kaplan-Meier estimates of the survival curve.

Usage

```

kmest(
  data,

```

```

rep = "",
stratum = "",
time = "time",
event = "event",
conftype = "log-log",
conflev = 0.95
)

```

Arguments

<code>data</code>	The input data frame that contains the following variables: <ul style="list-style-type: none"> • <code>rep</code>: The replication for by-group processing. • <code>stratum</code>: The stratum. • <code>time</code>: The possibly right-censored survival time. • <code>event</code>: The event indicator.
<code>rep</code>	The name(s) of the replication variable(s) in the input data.
<code>stratum</code>	The name(s) of the stratum variable(s) in the input data.
<code>time</code>	The name of the time variable in the input data.
<code>event</code>	The name of the event variable in the input data.
<code>conftype</code>	The type of the confidence interval. One of "none", "plain", "log", "log-log" (the default), or "arcsin". The arcsin option bases the intervals on $\text{asin}(\sqrt{\text{survival}})$.
<code>conflev</code>	The level of the two-sided confidence interval for the survival probabilities. Defaults to 0.95.

Value

A data frame with the following variables:

- `size`: The number of subjects in the stratum.
- `time`: The event time.
- `nrisk`: The number of subjects at risk.
- `nevent`: The number of subjects having the event.
- `survival`: The Kaplan-Meier estimate of the survival probability.
- `stderr`: The standard error of the estimated survival probability based on the Greenwood formula.
- `lower`: The lower bound of confidence interval if requested.
- `upper`: The upper bound of confidence interval if requested.
- `conflev`: The level of confidence interval if requested.
- `conftype`: The type of confidence interval if requested.
- `stratum`: The stratum.
- `rep`: The replication.

Author(s)

Kaifeng Lu, <kaifenglu@gmail.com>

Examples

```
kmemst(data = aml, stratum = "x", time = "time", event = "status")
```

liferegr

Parametric Regression Models for Failure Time Data

Description

Obtains the parameter estimates from parametric regression models with uncensored, right censored, left censored, or interval censored data.

Usage

```
liferegr(  
  data,  
  rep = "",  
  stratum = "",  
  time = "time",  
  time2 = "",  
  event = "event",  
  covariates = "",  
  weight = "",  
  offset = "",  
  id = "",  
  dist = "weibull",  
  robust = FALSE,  
  plci = FALSE,  
  alpha = 0.05  
)
```

Arguments

data	The input data frame that contains the following variables: <ul style="list-style-type: none">• rep: The replication for by-group processing.• stratum: The stratum.• time: The follow-up time for right censored data, or the left end of each interval for interval censored data.• time2: The right end of each interval for interval censored data.• event: The event indicator, 1=event, 0=no event.• covariates: The values of baseline covariates.• weight: The weight for each observation.
------	---

	<ul style="list-style-type: none"> • <code>offset</code>: The offset for each observation. • <code>id</code>: The optional subject ID to group the score residuals in computing the robust sandwich variance.
<code>rep</code>	The name(s) of the replication variable(s) in the input data.
<code>stratum</code>	The name(s) of the stratum variable(s) in the input data.
<code>time</code>	The name of the time variable or the left end of each interval for interval censored data in the input data.
<code>time2</code>	The name of the right end of each interval for interval censored data in the input data.
<code>event</code>	The name of the event variable in the input data for right censored data.
<code>covariates</code>	The vector of names of baseline covariates in the input data.
<code>weight</code>	The name of the weight variable in the input data.
<code>offset</code>	The name of the offset variable in the input data.
<code>id</code>	The name of the id variable in the input data.
<code>dist</code>	The assumed distribution for time to event. Options include "exponential", "weibull", "lognormal", and "loglogistic" to be modeled on the log-scale, and "normal" and "logistic" to be modeled on the original scale.
<code>robust</code>	Whether a robust sandwich variance estimate should be computed. In the presence of the id variable, the score residuals will be aggregated for each id when computing the robust sandwich variance estimate.
<code>plci</code>	Whether to obtain profile likelihood confidence interval.
<code>alpha</code>	The two-sided significance level.

Details

There are two ways to specify the model, one for right censored data through the time and event variables, and the other for interval censored data through the time (lower) and time2 (upper) variables. For the second form, we follow the convention used in SAS PROC LIFEREG:

- If lower is not missing, upper is not missing, and lower is equal to upper, then there is no censoring and the event occurred at time lower.
- If lower is not missing, upper is not missing, and lower < upper, then the event time is censored within the interval (lower, upper).
- If lower is missing, but upper is not missing, then upper will be used as the left censoring value.
- If lower is not missing, but upper is missing, then lower will be used as the right censoring value.
- If lower is not missing, upper is not missing, but lower > upper, or if both lower and upper are missing, then the observation will not be used.

Value

A list with the following components:

- `sumstat`: The data frame of summary statistics of model fit with the following variables:
 - `n`: The number of observations.
 - `nevents`: The number of events.
 - `loglik0`: The log-likelihood under null.
 - `loglik1`: The maximum log-likelihood.
 - `niter`: The number of Newton-Raphson iterations.
 - `dist`: The assumed distribution.
 - `p`: The number of parameters, including the intercept, regression coefficients associated with the covariates, and the log scale parameters for the strata.
 - `nvar`: The number of regression coefficients associated with the covariates (excluding the intercept).
 - `robust`: Whether the robust sandwich variance estimate is requested.
 - `rep`: The replication.
- `parest`: The data frame of parameter estimates with the following variables:
 - `param`: The name of the covariate for the parameter estimate.
 - `beta`: The parameter estimate.
 - `sebeta`: The standard error of parameter estimate.
 - `z`: The Wald test statistic for the parameter.
 - `expbeta`: The exponentiated parameter estimate.
 - `vbeta`: The covariance matrix for parameter estimates.
 - `lower`: The lower limit of confidence interval.
 - `upper`: The upper limit of confidence interval.
 - `p`: The p-value from the chi-square test.
 - `method`: The method to compute the confidence interval and p-value.
 - `sebeta_naive`: The naive standard error of parameter estimate if robust variance is requested.
 - `vbeta_naive`: The naive covariance matrix for parameter estimates if robust variance is requested.
 - `rep`: The replication.
- `p`: The number of parameters.
- `nvar`: The number of columns of the design matrix excluding the intercept.
- `param`: The parameter names.
- `beta`: The parameter estimate.
- `vbeta`: The covariance matrix for parameter estimates.
- `vbeta_naive`: The naive covariance matrix for parameter estimates.
- `terms`: The terms object.
- `xlevels`: A record of the levels of the factors used in fitting.
- `data`: The input data.

- rep: The name(s) of the replication variable(s).
- stratum: The name(s) of the stratum variable(s).
- time: The name of the time variable.
- time2: The name of the time2 variable.
- event: The name of the event variable.
- covariates: The names of baseline covariates.
- weight: The name of the weight variable.
- offset: The name of the offset variable.
- id: The name of the id variable.
- dist: The assumed distribution for time to event.
- robust: Whether a robust sandwich variance estimate should be computed.
- plci: Whether to obtain profile likelihood confidence interval.
- alpha: The two-sided significance level.

Author(s)

Kaifeng Lu, <kaifenglu@gmail.com>

References

John D. Kalbfleisch and Ross L. Prentice. The Statistical Analysis of Failure Time Data. Wiley: New York, 1980.

Examples

```
library(dplyr)

# right censored data
(fit1 <- liferegr(
  data = rawdata %>% mutate(treat = 1*(treatmentGroup == 1)),
  rep = "iterationNumber", stratum = "stratum",
  time = "timeUnderObservation", event = "event",
  covariates = "treat", dist = "weibull"))

# tobit regression for left censored data
(fit2 <- liferegr(
  data = tobin %>% mutate(time = ifelse(durable>0, durable, NA)),
  time = "time", time2 = "durable",
  covariates = c("age", "quant"), dist = "normal"))
```

Description

Obtains the parameter estimates from logistic regression models with binary data.

Usage

```
logisregr(  
  data,  
  rep = "",  
  event = "event",  
  covariates = "",  
  freq = "",  
  weight = "",  
  offset = "",  
  id = "",  
  robust = FALSE,  
  firth = FALSE,  
  flic = FALSE,  
  plci = FALSE,  
  alpha = 0.05  
)
```

Arguments

data	The input data frame that contains the following variables: <ul style="list-style-type: none">• rep: The replication for by-group processing.• event: The event indicator, 1=event, 0=no event.• covariates: The values of baseline covariates.• freq: The frequency for each observation.• weight: The weight for each observation.• offset: The offset for each observation.• id: The optional subject ID to group the score residuals in computing the robust sandwich variance.
rep	The name(s) of the replication variable(s) in the input data.
event	The name of the event variable in the input data.
covariates	The vector of names of baseline covariates in the input data.
freq	The name of the frequency variable in the input data. The frequencies must be the same for all observations within each cluster as indicated by the id. Thus freq is the cluster frequency.
weight	The name of the weight variable in the input data.
offset	The name of the offset variable in the input data.

<code>id</code>	The name of the id variable in the input data.
<code>robust</code>	Whether a robust sandwich variance estimate should be computed. In the presence of the id variable, the score residuals will be aggregated for each id when computing the robust sandwich variance estimate.
<code>firth</code>	Whether the firth's bias reducing penalized likelihood should be used. The default is FALSE.
<code>flic</code>	Whether to apply intercept correction to obtain more accurate predicted probabilities. The default is FALSE.
<code>plci</code>	Whether to obtain profile likelihood confidence interval.
<code>alpha</code>	The two-sided significance level.

Details

Fitting a logistic regression model using Firth's bias reduction method is equivalent to penalization of the log-likelihood by the Jeffreys prior. Firth's penalized log-likelihood is given by

$$l(\beta) + \frac{1}{2} \log(\det(I(\beta)))$$

and the components of the gradient $g(\beta)$ are computed as

$$g(\beta_j) + \frac{1}{2} \text{trace} \left(I(\beta)^{-1} \frac{\partial I(\beta)}{\partial \beta_j} \right)$$

The Hessian matrix is not modified by this penalty.

Firth's method reduces bias in maximum likelihood estimates of coefficients, but it introduces a bias toward one-half in the predicted probabilities.

A straightforward modification to Firth's logistic regression to achieve unbiased average predicted probabilities involves a post hoc adjustment of the intercept. This approach, known as Firth's logistic regression with intercept correction (FLIC), preserves the bias-corrected effect estimates. By excluding the intercept from penalization, it ensures that we don't sacrifice the accuracy of effect estimates to improve the predictions.

Value

A list with the following components:

- `sumstat`: The data frame of summary statistics of model fit with the following variables:
 - `n`: The number of subjects.
 - `nevents`: The number of events.
 - `loglik0`: The (penalized) log-likelihood under null.
 - `loglik1`: The maximum (penalized) log-likelihood.
 - `niter`: The number of Newton-Raphson iterations.
 - `p`: The number of parameters, including the intercept, and regression coefficients associated with the covariates.
 - `robust`: Whether a robust sandwich variance estimate should be computed.
 - `firth`: Whether the firth's penalized likelihood is used.

- flic: Whether to apply intercept correction.
- loglik0_unpenalized: The unpenalized log-likelihood under null.
- loglik1_unpenalized: The maximum unpenalized log-likelihood.
- rep: The replication.
- parest: The data frame of parameter estimates with the following variables:
 - param: The name of the covariate for the parameter estimate.
 - beta: The parameter estimate.
 - sebeta: The standard error of parameter estimate.
 - z: The Wald test statistic for the parameter.
 - expbeta: The exponentiated parameter estimate.
 - vbeta: The covariance matrix for parameter estimates.
 - lower: The lower limit of confidence interval.
 - upper: The upper limit of confidence interval.
 - p: The p-value from the chi-square test.
 - method: The method to compute the confidence interval and p-value.
 - sebeta_naive: The naive standard error of parameter estimate.
 - vbeta_naive: The naive covariance matrix of parameter estimates.
 - rep: The replication.
- fitted: The data frame with the following variables:
 - linear_predictors: The linear fit on the logit scale.
 - fitted_values: The fitted probabilities of having an event, obtained by transforming the linear predictors by the inverse of the logit link.
 - rep: The replication.
- p: The number of parameters.
- param: The parameter names.
- beta: The parameter estimate.
- vbeta: The covariance matrix for parameter estimates.
- vbeta_naive: The naive covariance matrix for parameter estimates.
- linear_predictors: The linear fit on the logit scale.
- fitted_values: The fitted probabilities of having an event.
- terms: The terms object.
- xlevels: A record of the levels of the factors used in fitting.
- data: The input data.
- rep: The name(s) of the replication variable(s).
- event: The name of the event variable.
- covariates: The names of baseline covariates.
- freq: The name of the freq variable.
- weight: The name of the weight variable.
- offset: The name of the offset variable.

- `id`: The name of the id variable.
- `robust`: Whether a robust sandwich variance estimate should be computed.
- `firth`: Whether to use the firth's bias reducing penalized likelihood.
- `flic`: Whether to apply intercept correction.
- `plci`: Whether to obtain profile likelihood confidence interval.
- `alpha`: The two-sided significance level.

Author(s)

Kaifeng Lu, <kaifenglu@gmail.com>

References

David Firth. Bias Reduction of Maximum Likelihood Estimates. *Biometrika* 1993; 80:27–38.

Georg Heinze and Michael Schemper. A solution to the problem of separation in logistic regression. *Statistics in Medicine* 2002;21:2409–2419.

Rainer Puhr, Georg Heinze, Mariana Nold, Lara Lusa, and Angelika Geroldinger. Firth's logistic regression with rare events: accurate effect estimates and predictions? *Statistics in Medicine* 2017; 36:2302-2317.

Examples

```
(fit1 <- logisregr(
  ingots, event = "NotReady", covariates = "Heat*Soak", freq = "Freq"))
```

Irttest

Log-Rank Test of Survival Curve Difference

Description

Obtains the log-rank test using the Fleming-Harrington family of weights.

Usage

```
Irttest(
  data,
  rep = "",
  stratum = "",
  treat = "treat",
  time = "time",
  event = "event",
  rho1 = 0,
  rho2 = 0
)
```


Arguments

data	The input data frame that contains the following variables: <ul style="list-style-type: none">• rep: The replication for by-group processing.• stratum: The stratum.• treat: The treatment.• time: The possibly right-censored survival time.• event: The event indicator.
rep	The name of the replication variable in the input data.
stratum	The name of the stratum variable in the input data.
treat	The name of the treatment variable in the input data.
time	The name of the time variable in the input data.
event	The name of the event variable in the input data.
rho1	The first parameter of the Fleming-Harrington family of weighted log-rank test. Defaults to 0 for conventional log-rank test.
rho2	The second parameter of the Fleming-Harrington family of weighted log-rank test. Defaults to 0 for conventional log-rank test.

Value

A data frame with the following variables:

- uscore: The numerator of the log-rank test statistic.
- vscore: The variance of the log-rank score test statistic.
- logRankZ: The Z-statistic value.
- logRankPValue: The one-sided p-value.
- rho1: The first parameter of the Fleming-Harrington weights.
- rho2: The second parameter of the Fleming-Harrington weights.
- rep: The replication.

Author(s)

Kaifeng Lu, <kaifenglu@gmail.com>

Examples

```
df <- Irtest(data = rawdata, rep = "iterationNumber",
            stratum = "stratum", treat = "treatmentGroup",
            time = "timeUnderObservation", event = "event",
            rho1 = 0.5, rho2 = 0)
head(df)
```

Description

Computes the B-spline basis matrix for a natural cubic spline.

Usage

```
nscpp(  
  x = NA_real_,  
  df = NA_integer_,  
  knots = NA_real_,  
  intercept = 0L,  
  boundary_knots = NA_real_  
)
```

Arguments

- | | |
|----------------|--|
| x | A numeric vector representing the predictor variable. Missing values are allowed. |
| df | Degrees of freedom, specifying the number of columns in the basis matrix. If df is provided, the function selects $df - 1 - \text{intercept}$ internal knots based on appropriate quantiles of x, ignoring any missing values. |
| knots | A numeric vector specifying the internal breakpoints that define the spline. If provided, the number of degrees of freedom will be determined by the length of knots. |
| intercept | A logical value indicating whether to include an intercept in the basis. The default is FALSE. |
| boundary_knots | A numeric vector of length 2 specifying the boundary points where the natural boundary conditions are applied and the B-spline basis is anchored. If not supplied, the default is the range of non-missing values in x. |

Value

A matrix with dimensions $c(\text{length}(x), df)$, where df is either provided directly or computed as $\text{length}(\text{knots}) + 1 + \text{intercept}$ when knots are supplied. The matrix contains attributes that correspond to the arguments passed to the nscpp function.

Author(s)

Kaifeng Lu, <kaifenglu@gmail.com>

Examples

```
nscpp(women$height, df = 5)
```

phregr

Proportional Hazards Regression Models

Description

Obtains the hazard ratio estimates from the proportional hazards regression model with right censored or counting process data.

Usage

```
phregr(  
  data,  
  rep = "",  
  stratum = "",  
  time = "time",  
  time2 = "",  
  event = "event",  
  covariates = "",  
  weight = "",  
  offset = "",  
  id = "",  
  ties = "efron",  
  robust = FALSE,  
  est_basehaz = TRUE,  
  est_resid = TRUE,  
  firth = FALSE,  
  plci = FALSE,  
  alpha = 0.05  
)
```

Arguments

data

The input data frame that contains the following variables:

- rep: The replication for by-group processing.
- stratum: The stratum.
- time: The follow-up time for right censored data, or the left end of each interval for counting process data.
- time2: The right end of each interval for counting process data. Intervals are assumed to be open on the left and closed on the right, and event indicates whether an event occurred at the right end of each interval.
- event: The event indicator, 1=event, 0=no event.
- covariates: The values of baseline covariates (and time-dependent covariates in each interval for counting process data).
- weight: The weight for each observation.
- offset: The offset for each observation.

	<ul style="list-style-type: none"> • <code>id</code>: The optional subject ID for counting process data with time-dependent covariates.
<code>rep</code>	The name(s) of the replication variable(s) in the input data.
<code>stratum</code>	The name(s) of the stratum variable(s) in the input data.
<code>time</code>	The name of the time variable or the left end of each interval for counting process data in the input data.
<code>time2</code>	The name of the right end of each interval for counting process data in the input data.
<code>event</code>	The name of the event variable in the input data.
<code>covariates</code>	The vector of names of baseline and time-dependent covariates in the input data.
<code>weight</code>	The name of the weight variable in the input data.
<code>offset</code>	The name of the offset variable in the input data.
<code>id</code>	The name of the id variable in the input data.
<code>ties</code>	The method for handling ties, either "breslow" or "efron" (default).
<code>robust</code>	Whether a robust sandwich variance estimate should be computed. In the presence of the <code>id</code> variable, the score residuals will be aggregated for each <code>id</code> when computing the robust sandwich variance estimate.
<code>est_basehaz</code>	Whether to estimate the baseline hazards. Defaults to TRUE.
<code>est_resid</code>	Whether to estimate the martingale residuals. Defaults to TRUE.
<code>firth</code>	Whether to use Firth's penalized likelihood method. Defaults to FALSE.
<code>plci</code>	Whether to obtain profile likelihood confidence interval.
<code>alpha</code>	The two-sided significance level.

Value

A list with the following components:

- `sumstat`: The data frame of summary statistics of model fit with the following variables:
 - `n`: The number of observations.
 - `nevents`: The number of events.
 - `loglik0`: The (penalized) log-likelihood under null.
 - `loglik1`: The maximum (penalized) log-likelihood.
 - `scoretest`: The score test statistic.
 - `niter`: The number of Newton-Raphson iterations.
 - `ties`: The method for handling ties, either "breslow" or "efron".
 - `p`: The number of columns of the Cox model design matrix.
 - `robust`: Whether to use the robust variance estimate.
 - `firth`: Whether to use Firth's penalized likelihood method.
 - `loglik0_unpenalized`: The unpenalized log-likelihood under null.
 - `loglik1_unpenalized`: The maximum unpenalized log-likelihood.
 - `rep`: The replication.
- `parest`: The data frame of parameter estimates with the following variables:

- param: The name of the covariate for the parameter estimate.
- beta: The log hazard ratio estimate.
- sebeta: The standard error of log hazard ratio estimate.
- z: The Wald test statistic for log hazard ratio.
- expbeta: The hazard ratio estimate.
- vbeta: The covariance matrix for parameter estimates.
- lower: The lower limit of confidence interval.
- upper: The upper limit of confidence interval.
- p: The p-value from the chi-square test.
- method: The method to compute the confidence interval and p-value.
- sebeta_naive: The naive standard error of log hazard ratio estimate if robust variance is requested.
- vbeta_naive: The naive covariance matrix for parameter estimates if robust variance is requested.
- rep: The replication.
- basehaz: The data frame of baseline hazards with the following variables (if est_basehaz is TRUE):
 - time: The observed event time.
 - nrisk: The number of patients at risk at the time point.
 - nevent: The number of events at the time point.
 - haz: The baseline hazard at the time point.
 - varhaz: The variance of the baseline hazard at the time point assuming the parameter beta is known.
 - gradhaz: The gradient of the baseline hazard with respect to beta at the time point (in the presence of covariates).
 - stratum: The stratum.
 - rep: The replication.
- residuals: The martingale residuals.
- p: The number of parameters.
- param: The parameter names.
- beta: The parameter estimate.
- vbeta: The covariance matrix for parameter estimates.
- vbeta_naive: The naive covariance matrix for parameter estimates.
- terms: The terms object.
- xlevels: A record of the levels of the factors used in fitting.
- data: The input data.
- rep: The name(s) of the replication variable(s).
- stratum: The name(s) of the stratum variable(s).
- time: The name of the time variable.
- time2: The name of the time2 variable.
- event: The name of the event variable.

- `covariates`: The names of baseline covariates.
- `weight`: The name of the weight variable.
- `offset`: The name of the offset variable.
- `id`: The name of the id variable.
- `ties`: The method for handling ties.
- `robust`: Whether a robust sandwich variance estimate should be computed.
- `est_basehaz`: Whether to estimate the baseline hazards.
- `est_resid`: Whether to estimate the martingale residuals.
- `firth`: Whether to use Firth's penalized likelihood method.
- `plci`: Whether to obtain profile likelihood confidence interval.
- `alpha`: The two-sided significance level.

Author(s)

Kaifeng Lu, <kaifenglu@gmail.com>

References

Per K. Anderson and Richard D. Gill. Cox's regression model for counting processes, a large sample study. *Annals of Statistics* 1982; 10:1100-1120.

Terry M. Therneau and Patricia M. Grambsch. *Modeling Survival Data: Extending the Cox Model*. Springer-Verlag, 2000.

Examples

```
library(dplyr)

# Example 1 with right-censored data
(fit1 <- phregr(
  data = rawdata %>% mutate(treat = 1*(treatmentGroup == 1)),
  rep = "iterationNumber", stratum = "stratum",
  time = "timeUnderObservation", event = "event",
  covariates = "treat", est_basehaz = FALSE, est_resid = FALSE))

# Example 2 with counting process data and robust variance estimate
(fit2 <- phregr(
  data = heart %>% mutate(rx = as.numeric(transplant) - 1),
  time = "start", time2 = "stop", event = "event",
  covariates = c("rx", "age"), id = "id",
  robust = TRUE, est_basehaz = TRUE, est_resid = TRUE))
```

qr/cpp

*QR Decomposition of a Matrix***Description**

Computes the QR decomposition of a matrix.

Usage

```
qr/cpp(x, tol = 1e-12)
```

Arguments

`x` A numeric matrix whose QR decomposition is to be computed.
`tol` The tolerance for detecting linear dependencies in the columns of `x`.

Details

This function performs Householder QR with column pivoting: Given an m -by- n matrix A with $m \geq n$, the following algorithm computes $r = \text{rank}(A)$ and the factorization $Q^T A P$ equal to

$$\begin{array}{ccc|cc} & R_{11} & R_{12} & & r \\ & \mathbf{0} & \mathbf{0} & & m-r \\ r & & & n-r & \end{array}$$

with $Q = H_1 \cdots H_r$ and $P = P_1 \cdots P_r$. The upper triangular part of A is overwritten by the upper triangular part of R and components $(j+1) : m$ of the j th Householder vector are stored in $A((j+1) : m, j)$. The permutation P is encoded in an integer vector `pivot`.

Value

A list with the following components:

- `qr`: A matrix with the same dimensions as `x`. The upper triangle contains the R of the decomposition and the lower triangle contains Householder vectors (stored in compact form).
- `rank`: The rank of `x` as computed by the decomposition.
- `pivot`: The column permutation for the pivoting strategy used during the decomposition.
- `Q`: The complete m -by- m orthogonal matrix Q .
- `R`: The complete m -by- n upper triangular matrix R .

Author(s)

Kaifeng Lu, <kaifenglu@gmail.com>

References

Gene N. Golub and Charles F. Van Loan. Matrix Computations, second edition. Baltimore, Maryland: The John Hopkins University Press, 1989, p.235.

Examples

```
hilbert <- function(n) { i <- 1:n; 1 / outer(i - 1, i, `+`) }  
h9 <- hilbert(9)  
qrcpp(h9)
```

rawdata

A simulated time-to-event data set with 10 replications

Description

A simulated data set with stratification and delayed treatment effect:

`iterationNumber` The iteration number

`arrivalTime` The enrollment time for the subject

`stratum` The stratum for the subject

`treatmentGroup` The treatment group for the subject

`timeUnderObservation` The time under observation since randomization

`event` Whether the subject experienced the event

`dropoutEvent` Whether the subject dropped out

Usage

```
rawdata
```

Format

An object of class `data.frame` with 4910 rows and 7 columns.

residuals_phregr	<i>Residuals for Proportional Hazards Regression Models</i>
------------------	---

Description

Obtains the martingale, deviance, score, or Schoenfeld residuals for a proportional hazards regression model.

Usage

```
residuals_phregr(
  fit_phregr,
  type = c("martingale", "deviance", "score", "schoenfeld", "dfbeta", "dfbetas",
    "scaledsch"),
  collapse = FALSE,
  weighted = (type %in% c("dfbeta", "dfbetas"))
)
```

Arguments

fit_phregr	The output from the phregr call.
type	The type of residuals desired, with options including "martingale", "deviance", "score", "schoenfeld", "dfbeta", "dfbetas", and "scaledsch".
collapse	Whether to collapse the residuals by id. This is not applicable for Schoenfeld type residuals.
weighted	Whether to compute weighted residuals.

Details

For score and Schoenfeld type residuals, the proportional hazards model must include at least one covariate. The algorithms for deviance, dfbeta, dfbetas, and scaledsch residuals follow the residuals.coxph function in the survival package.

Value

For martingale and deviance residuals, the result is a vector with one element corresponding to each subject (without collapse). For score residuals, the result is a matrix where each row represents a subject and each column corresponds to a variable. The row order aligns with the input data used in the original fit. For Schoenfeld residuals, the result is a matrix with one row for each event and one column per variable. These rows are sorted by time within strata, with the attributes stratum and time included.

Score residuals represent each individual's contribution to the score vector. Two commonly used transformations of this are dfbeta, which represents the approximate change in the coefficient vector if the observation is excluded, and dfbetas, which gives the approximate change in the coefficients scaled by the standard error of the coefficients.

Author(s)

Kaifeng Lu, <kaifenglu@gmail.com>

References

Terry M. Therneau, Patricia M. Grambsch, and Thomas M. Fleming. Martingale based residuals for survival models. *Biometrika* 1990; 77:147-160.

Patricia M. Grambsch and Terry M. Therneau. Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika* 1994; 81:515-26.

Examples

```
library(dplyr)

# Example 1 with right-censored data
fit1 <- phregr(data = rawdata %>% filter(iterationNumber == 1) %>%
  mutate(treat = 1*(treatmentGroup == 1)),
  stratum = "stratum",
  time = "timeUnderObservation", event = "event",
  covariates = "treat")

ressco <- residuals_phregr(fit1, type = "score")

# Example 2 with counting process data
fit2 <- phregr(data = heart %>% mutate(rx = as.numeric(transplant) - 1),
  time = "start", time2 = "stop", event = "event",
  covariates = c("rx", "age"), id = "id", robust = TRUE)

resssch <- residuals_phregr(fit2, type = "scaledsch")
```

 rmdiff

Estimate of Restricted Mean Survival Time Difference

Description

Obtains the estimate of restricted mean survival time difference between two treatment groups.

Usage

```
rmdiff(
  data,
  rep = "",
  stratum = "",
  treat = "treat",
  time = "time",
  event = "event",
  milestone = NA_real_,
```

```

rmstDiffH0 = 0,
conflv = 0.95,
biascorrection = 0L
)

```

Arguments

data	The input data frame that contains the following variables: <ul style="list-style-type: none"> • rep: The replication for by-group processing. • stratum: The stratum. • treat: The treatment. • time: The possibly right-censored survival time. • event: The event indicator.
rep	The name of the replication variable in the input data.
stratum	The name of the stratum variable in the input data.
treat	The name of the treatment variable in the input data.
time	The name of the time variable in the input data.
event	The name of the event variable in the input data.
milestone	The milestone time at which to calculate the restricted mean survival time.
rmstDiffH0	The difference in restricted mean survival times under the null hypothesis. Defaults to 0 for superiority test.
conflv	The level of the two-sided confidence interval for the difference in restricted mean survival times. Defaults to 0.95.
biascorrection	Whether to apply bias correction for the variance estimate of individual restricted mean survival times. Defaults to no bias correction.

Value

A data frame with the following variables:

- rep: The replication number.
- milestone: The milestone time relative to randomization.
- rmstDiffH0: The difference in restricted mean survival times under the null hypothesis.
- rmst1: The estimated restricted mean survival time for the treatment group.
- rmst2: The estimated restricted mean survival time for the control group.
- rmstDiff: The estimated difference in restricted mean survival times.
- vrmst1: The variance for rmst1.
- vrmst2: The variance for rmst2.
- vrmstDiff: The variance for rmstDiff.
- rmstDiffZ: The Z-statistic value.
- rmstDiffPValue: The one-sided p-value.
- lower: The lower bound of confidence interval.

- upper: The upper bound of confidence interval.
- conflev: The level of confidence interval.
- biascorrection: Whether to apply bias correction for the variance estimate of individual restricted mean survival times.

Author(s)

Kaifeng Lu, <kaifenglu@gmail.com>

Examples

```
df <- rmdiff(data = rawdata, rep = "iterationNumber",
             stratum = "stratum", treat = "treatmentGroup",
             time = "timeUnderObservation", event = "event",
             milestone = 12)
head(df)
```

rmest

Estimate of Restricted Mean Survival Time

Description

Obtains the estimate of restricted means survival time for each stratum.

Usage

```
rmest(
  data,
  rep = "",
  stratum = "",
  time = "time",
  event = "event",
  milestone = NA_real_,
  conflev = 0.95,
  biascorrection = 0L
)
```

Arguments

data	The input data frame that contains the following variables: <ul style="list-style-type: none"> • rep: The replication for by-group processing. • stratum: The stratum. • time: The possibly right-censored survival time. • event: The event indicator.
rep	The name of the replication variable in the input data.

stratum	The name of the stratum variable in the input data.
time	The name of the time variable in the input data.
event	The name of the event variable in the input data.
milestone	The milestone time at which to calculate the restricted mean survival time.
conflev	The level of the two-sided confidence interval for the survival probabilities. Defaults to 0.95.
biascorrection	Whether to apply bias correction for the variance estimate. Defaults to no bias correction.

Value

A data frame with the following variables:

- rep: The replication.
- stratum: The stratum variable.
- size: The number of subjects in the stratum.
- milestone: The milestone time relative to randomization.
- rmst: The estimate of restricted mean survival time.
- stderr: The standard error of the estimated rmst.
- lower: The lower bound of confidence interval if requested.
- upper: The upper bound of confidence interval if requested.
- conflev: The level of confidence interval if requested.
- biascorrection: Whether to apply bias correction for the variance estimate.

Author(s)

Kaifeng Lu, <kaifenglu@gmail.com>

Examples

```
rrest(data = aml, stratum = "x",
      time = "time", event = "status", milestone = 24)
```

rpsftm

Rank Preserving Structural Failure Time Model (RPSFTM) for Treatment Switching

Description

Obtains the causal parameter estimate from the log-rank test and the hazard ratio estimate from the Cox model to account for treatment switching.

Usage

```
rpsftm(
  data,
  stratum = "",
  time = "time",
  event = "event",
  treat = "treat",
  rx = "rx",
  censor_time = "censor_time",
  base_cov = "",
  low_psi = -1,
  hi_psi = 1,
  n_eval_z = 100,
  treat_modifier = 1,
  recensor = TRUE,
  admin_recensor_only = FALSE,
  autoswitch = TRUE,
  gridsearch = FALSE,
  alpha = 0.05,
  ties = "efron",
  tol = 1e-06,
  boot = FALSE,
  n_boot = 1000,
  seed = NA
)
```

Arguments

data	The input data frame that contains the following variables: <ul style="list-style-type: none"> • stratum: The stratum. • time: The survival time for right censored data. • event: The event indicator, 1=event, 0=no event. • treat: The randomized treatment indicator, 1=treatment, 0=control. • rx: The proportion of time on active treatment. • censor_time: The administrative censoring time. It should be provided for all subjects including those who had events. • base_cov: The baseline covariates (excluding treat).
stratum	The name(s) of the stratum variable(s) in the input data.
time	The name of the time variable in the input data.
event	The name of the event variable in the input data.
treat	The name of the treatment variable in the input data.
rx	The name of the rx variable in the input data.
censor_time	The name of the censor_time variable in the input data.
base_cov	The vector of names of baseline covariates (excluding treat) in the input data.
low_psi	The lower limit of the causal parameter.

hi_psi	The upper limit of the causal parameter.
n_eval_z	The number of points between low_psi and hi_psi (inclusive) at which to evaluate the log-rank Z-statistics.
treat_modifier	The optional sensitivity parameter for the constant treatment effect assumption.
recensor	Whether to apply recensoring to counter-factual survival times. Defaults to TRUE.
admin_recensor_only	Whether to apply recensoring to administrative censoring time only. Defaults to FALSE, in which case, recensoring will be applied to the actual censoring time for dropouts.
autoswitch	Whether to exclude recensoring for treatment arms with no switching. Defaults to TRUE.
gridsearch	Whether to use grid search to estimate the causal parameter psi. Defaults to FALSE, in which case, a root finding algorithm will be used.
alpha	The significance level to calculate confidence intervals.
ties	The method for handling ties in the Cox model, either "breslow" or "efron" (default).
tol	The desired accuracy (convergence tolerance) for psi.
boot	Whether to use bootstrap to obtain the confidence interval for hazard ratio. Defaults to FALSE, in which case, the confidence interval will be constructed to match the log-rank test p-value.
n_boot	The number of bootstrap samples.
seed	The seed to reproduce the bootstrap results. The seed from the environment will be used if left unspecified.

Details

We use the following steps to obtain the hazard ratio estimate and confidence interval had there been no treatment switching:

- Use RPSFTM to estimate the causal parameter ψ based on the log-rank test for counter-factual untreated survival times for both arms: $U_{i,\psi} = T_{C_i} + e^{\psi}T_{E_i}$.
- Fit the Cox proportional hazards model to the observed survival times on the treatment arm and the counter-factual untreated survival times on the control arm to obtain the hazard ratio estimate.
- Use either the log-rank test p-value for the treatment policy strategy or bootstrap to construct the confidence interval for hazard ratio.

Value

A list with the following components:

- psi: The estimated causal parameter.
- psi_CI: The confidence interval for psi.

- `psi_CI_type`: The type of confidence interval for `psi`, i.e., "grid search", "root finding", or "bootstrap".
- `Sstar`: A data frame containing the counter-factual untreated survival times and the event indicators.
- `kmstar`: A data frame containing the Kaplan-Meier estimates based on the counter-factual untreated survival times by treatment arm.
- `eval_z`: A data frame containing the log-rank test Z-statistics evaluated at a sequence of `psi` values. Used to plot and check if the range of `psi` values to search for the solution and limits of confidence interval of `psi` need be modified.
- `logrank_pvalue`: The two-sided p-value of the log-rank test based on the treatment policy strategy.
- `cox_pvalue`: The two-sided p-value for treatment effect based on the Cox model.
- `hr`: The estimated hazard ratio from the Cox model.
- `hr_CI`: The confidence interval for hazard ratio.
- `hr_CI_type`: The type of confidence interval for hazard ratio, either "log-rank p-value" or "bootstrap".
- `settings`: A list with the following components:
 - `low_psi`: The lower limit of the causal parameter.
 - `hi_psi`: The upper limit of the causal parameter.
 - `n_eval_z`: The number of points between `low_psi` and `hi_psi` at which to evaluate the log-rank Z-statistics.
 - `treat_modifier`: The sensitivity parameter for the constant treatment effect assumption.
 - `recensor`: Whether to apply recensoring to counter-factual survival times.
 - `admin_recensor_only`: Whether to apply recensoring to administrative censoring time only.
 - `autoswitch`: Whether to exclude recensoring for treatment arms with no switching.
 - `gridsearch`: Whether to use grid search to estimate the causal parameter `psi`.
 - `alpha`: The significance level to calculate confidence intervals.
 - `ties`: The method for handling ties in the Cox model.
 - `tol`: The desired accuracy (convergence tolerance).
 - `boot`: Whether to use bootstrap to obtain the confidence interval for hazard ratio.
 - `n_boot`: The number of bootstrap samples.
 - `seed`: The seed to reproduce the simulation results.
- `hr_boots`: The bootstrap hazard ratio estimates if `boot` is TRUE.
- `psi_boots`: The bootstrap `psi` estimates if `boot` is TRUE.

Author(s)

Kaifeng Lu, <kaifengl@gmail.com>

References

James M. Robins and Anastasios A. Tsiatis. Correcting for non-compliance in randomized trials using rank preserving structural failure time models. *Communications in Statistics*. 1991;20(8):2609-2631.

Ian R. White, Abdel G. Babiker, Sarah Walker, and Janet H. Darbyshire. Randomization-based methods for correcting for treatment changes: Examples from the CONCORDE trial. *Statistics in Medicine*. 1999;18:2617-2634.

Examples

```
library(dplyr)

# Example 1: one-way treatment switching (control to active)

data <- immdef %>% mutate(rx = 1-xoyrs/progyrs)

fit1 <- rpsftm(
  data, time = "progyrs", event = "prog", treat = "imm",
  rx = "rx", censor_time = "censyrs", boot = FALSE)

c(fit1$hr, fit1$hr_CI)

# Example 2: two-way treatment switching (illustration only)

# the eventual survival time
shilong1 <- shilong %>%
  arrange(bras.f, id, tstop) %>%
  group_by(bras.f, id) %>%
  filter(row_number() == n()) %>%
  select(-c("ps", "ttc", "tran"))

shilong2 <- shilong1 %>%
  mutate(rx = ifelse(co, ifelse(bras.f == "MTA", dco/ady,
    1 - dco/ady),
    ifelse(bras.f == "MTA", 1, 0)))

fit2 <- rpsftm(
  shilong2, time = "tstop", event = "event",
  treat = "bras.f", rx = "rx", censor_time = "dcut",
  base_cov = c("agerand", "sex.f", "tt_Lnum", "rmh_alea.c",
    "pathway.f"),
  low_psi = -3, hi_psi = 3, boot = FALSE)

c(fit2$hr, fit2$hr_CI)
```

Description

This data set deals with urinary tract infection in sexually active college women, along with covariate information on age and contraceptive use. The variables are all binary and coded in 1 (condition is present) and 0 (condition is absent).

Usage

```
sexagg
```

Format

An object of class `data.frame` with 36 rows and 9 columns.

Details

`case` urinary tract infection, the study outcome variable

`age` ≥ 24 years

`dia` use of diaphragm

`oc` use of oral contraceptive

`vic` use of condom

`vicl` use of lubricated condom

`vis` use of spermicide

shilong

The randomized clinical trial SHIVA data in long format from the ipcswitch package

Description

The original SHIdat data set contains an anonymized excerpt of data from the SHIVA01 trial. This was the first randomized clinical trial that aimed at comparing molecularly targeted therapy based on tumor profiling (MTA) versus conventional therapy (CT) for advanced cancer. Patients were randomly assigned to receive the active or control treatment and may switch to the other arm or subsequent anti-cancer therapy upon disease progression. The restructured data is in the long format.

`id` The patient's identifier

`tstart` The start of the time interval

`tstop` The end of the time interval

`event` Whether the patient died at the end of the interval

`agerand` The patient's age (in years) at randomization

`sex.f` The patients' gender, either Male or Female

`tt_Lnum` The number of previous lines of treatment

rmh_alea.c The Royal Marsden Hospital score segregated into two categories
 pathway.f The molecular pathway altered (the hormone receptors pathway, the PI3K/ AKT/mTOR
 pathway, and the RAF/MEK pathway)
 bras.f The patient's randomized arm, either MTA or CT
 ps The ECOG performance status
 ttc The presence of concomitant treatments
 tran The use of platelet transfusions
 dpd The relative day of a potential progression
 dco The relative day of treatment switching
 ady The relative day of the latest news
 dcut The relative day of administrative cutoff
 pd Whether the patient had disease progression
 co Whether the patient switched treatment

Usage

shilong

Format

An object of class `data.frame` with 602 rows and 19 columns.

six	<i>The repeated measures data from the "Six Cities" study of the health effects of air pollution (Ware et al. 1984).</i>
-----	--

Description

The data analyzed are the 16 selected cases in Lipsitz et al. (1994). The binary response is the wheezing status of 16 children at ages 9, 10, 11, and 12 years. A value of 1 of wheezing status indicates the occurrence of wheezing. The explanatory variables city of residence, age, and maternal smoking status at the particular age.

Usage

six

Format

An object of class `tbl_df` (inherits from `tbl`, `data.frame`) with 64 rows and 6 columns.

Details

case case id
 city city of residence
 age age of the child
 smoke maternal smoking status
 wheeze wheezing status

splineDesigncpp *B-Spline Design Matrix*

Description

Computes the design matrix for B-splines based on the specified knots and evaluated at the values in x .

Usage

```
splineDesigncpp(
  knots = NA_real_,
  x = NA_real_,
  ord = 4L,
  derivs = as.integer(c(0))
)
```

Arguments

knots	A numeric vector specifying the positions of the knots, including both boundary and internal knots.
x	A numeric vector of values where the B-spline functions or their derivatives will be evaluated. The values of x must lie within the range of the "inner" knots, i.e., between $\text{knots}[\text{ord}]$ and $\text{knots}[\text{length}(\text{knots}) - (\text{ord} - 1)]$.
ord	A positive integer indicating the order of the B-spline. This corresponds to the number of coefficients in each piecewise polynomial segment, where $\text{ord} = \text{degree} + 1$.
derivs	An integer vector specifying the order of derivatives to be evaluated at the corresponding x values. Each value must be between 0 and $\text{ord} - 1$, and the vector is conceptually recycled to match the length of x . The default is 0, meaning the B-spline functions themselves are evaluated.

Value

A matrix with dimensions $c(\text{length}(x), \text{length}(\text{knots}) - \text{ord})$. Each row corresponds to a value in x and contains the coefficients of the B-splines, or the specified derivatives, as defined by the knots and evaluated at that particular value of x . The total number of B-splines is $\text{length}(\text{knots}) - \text{ord}$, with each B-spline defined by a set of ord consecutive knots.

Author(s)

Kaifeng Lu, <kaifenglu@gmail.com>

Examples

```
splineDesigncpp(knots = 1:10, x = 4:7)
splineDesigncpp(knots = 1:10, x = 4:7, derivs = 1)
```

survfit_phreg

Survival Curve for Proportional Hazards Regression Models

Description

Obtains the predicted survivor function for a proportional hazards regression model.

Usage

```
survfit_phreg(  
  fit_phreg,  
  newdata,  
  sefit = TRUE,  
  conftype = "log-log",  
  conflev = 0.95  
)
```

Arguments

fit_phreg	The output from the phreg call.
newdata	A data frame with the same variable names as those that appear in the phreg call. For right-censored data, one curve is produced per row to represent a cohort whose covariates correspond to the values in newdata. For counting-process data, one curve is produced per id in newdata to present the survival curve along the path of time-dependent covariates at the observed event times in the data used to fit phreg.
sefit	Whether to compute the standard error of the survival estimates.
conftype	The type of the confidence interval. One of "none", "plain", "log", "log-log" (the default), or "arcsin". The arcsin option bases the intervals on $\text{asin}(\sqrt{\text{surv}})$.
conflev	The level of the two-sided confidence interval for the survival probabilities. Defaults to 0.95.

Details

If newdata is not provided and there is no covariate, survival curves based on the basehaz data frame will be produced.

Value

A data frame with the following variables:

- `id`: The id of the subject for counting-process data with time-dependent covariates.
- `time`: The observed times in the data used to fit `phreg`.
- `nrisk`: The number of patients at risk at the time point in the data used to fit `phreg`.
- `nevent`: The number of patients having event at the time point in the data used to fit `phreg`.
- `cumhaz`: The cumulative hazard at the time point.
- `surv`: The estimated survival probability at the time point.
- `sesurv`: The standard error of the estimated survival probability.
- `lower`: The lower confidence limit for survival probability.
- `upper`: The upper confidence limit for survival probability.
- `confl`: The level of the two-sided confidence interval.
- `conftype`: The type of the confidence interval.
- `covariates`: The values of covariates based on `newdata`.
- `stratum`: The stratum of the subject.

Author(s)

Kaifeng Lu, <kaifenglu@gmail.com>

References

Terry M. Therneau and Patricia M. Grambsch. *Modeling Survival Data: Extending the Cox Model*. Springer-Verlag, 2000.

Examples

```
library(dplyr)

# Example 1 with right-censored data
fit1 <- phregr(data = rawdata %>% filter(iterationNumber == 1) %>%
  mutate(treat = 1*(treatmentGroup == 1)),
  stratum = "stratum",
  time = "timeUnderObservation", event = "event",
  covariates = "treat")

surv1 <- survfit_phreg(fit1,
  newdata = data.frame(
    stratum = as.integer(c(1,1,2,2)),
    treat = c(1,0,1,0)))

# Example 2 with counting process data and robust variance estimate
fit2 <- phregr(data = heart %>% mutate(rx = as.numeric(transplant) - 1),
  time = "start", time2 = "stop", event = "event",
  covariates = c("rx", "age"), id = "id", robust = TRUE)
```

```
surv2 <- survfit_phregr(fit2,
  newdata = data.frame(
    id = c(4,4,11,11),
    age = c(-7.737,-7.737,-0.019,-0.019),
    start = c(0,36,0,26),
    stop = c(36,39,26,153),
    rx = c(0,1,0,1)))
```

tobin

Tobin's tobit data from the survival package

Description

Data from Tobin's original paper.

durable Durable goods purchase

age Age in years

quant Liquidity ratio (x 1000)

Usage

```
tobin
```

Format

An object of class `data.frame` with 20 rows and 3 columns.

tsegest

The Two-Stage Estimation (TSE) Method Using g-estimation for Treatment Switching

Description

Obtains the causal parameter estimate of the logistic regression switching model and the hazard ratio estimate of the Cox model to account for treatment switching.

Usage

```
tsegest(
  data,
  id = "id",
  stratum = "",
  tstart = "tstart",
  tstop = "tstop",
  event = "event",
```

```

treat = "treat",
censor_time = "censor_time",
pd = "pd",
pd_time = "pd_time",
swtrt = "swtrt",
swtrt_time = "swtrt_time",
swtrt_time_upper = "",
base_cov = "",
conf_cov = "",
low_psi = -3,
hi_psi = 3,
strata_main_effect_only = TRUE,
firth = FALSE,
flic = FALSE,
recensor = TRUE,
admin_recensor_only = FALSE,
swtrt_control_only = TRUE,
alpha = 0.05,
ties = "efron",
tol = 1e-06,
boot = TRUE,
n_boot = 1000,
seed = NA
)

```

Arguments

data	<p>The input data frame that contains the following variables:</p> <ul style="list-style-type: none"> • id: The id to identify observations belonging to the same subject for counting process data with time-dependent covariates. • stratum: The stratum. • tstart: The starting time of the time interval for counting-process data with time-dependent covariates. • tstop: The stopping time of the time interval for counting-process data with time-dependent covariates. • event: The event indicator, 1=event, 0=no event. • treat: The randomized treatment indicator, 1=treatment, 0=control. • censor_time: The administrative censoring time. It should be provided for all subjects including those who had events. • pd: The disease progression indicator, 1=PD, 0=no PD. • pd_time: The time from randomization to PD. • swtrt: The treatment switch indicator, 1=switch, 0=no switch. • swtrt_time: The time from randomization to treatment switch. • swtrt_time_upper: The upper bound of treatment switching time. • base_cov: The values of baseline covariates (excluding treat). • conf_cov: The values of confounding variables for predicting treatment switching (excluding treat).
------	--

id	The name of the id variable in the input data.
stratum	The name(s) of the stratum variable(s) in the input data.
tstart	The name of the tstart variable in the input data.
tstop	The name of tstop variable in the input data.
event	The name of the event variable in the input data.
treat	The name of the treatment variable in the input data.
sensor_time	The name of the sensor_time variable in the input data.
pd	The name of the pd variable in the input data.
pd_time	The name of the pd_time variable in the input data.
swtrt	The name of the swtrt variable in the input data.
swtrt_time	The name of the swtrt_time variable in the input data.
swtrt_time_upper	The name of the swtrt_time_upper variable in the input data.
base_cov	The vector of names of base_cov variables (excluding treat) in the input data for the Cox model.
conf_cov	The vector of the names of conf_cov variables (excluding treat) in the input data for the logistic regression switching model.
low_psi	The lower limit of the causal parameter.
hi_psi	The upper limit of the causal parameter.
strata_main_effect_only	Whether to only include the strata main effects in the logistic regression switching model. Defaults to TRUE, otherwise all possible strata combinations will be considered in the switching model.
firth	Whether the firth's bias reducing penalized likelihood should be used. The default is FALSE.
flic	Whether to apply intercept correction to obtain more accurate predicted probabilities. The default is FALSE.
recensor	Whether to apply recensoring to counter-factual survival times. Defaults to TRUE.
admin_recensor_only	Whether to apply recensoring to administrative censoring time only. Defaults to FALSE, in which case, recensoring will be applied to the actual censoring time for dropouts.
swtrt_control_only	Whether treatment switching occurred only in the control group.
alpha	The significance level to calculate confidence intervals.
ties	The method for handling ties in the Cox model, either "breslow" or "efron" (default).
tol	The desired accuracy (convergence tolerance) for psi.
boot	Whether to use bootstrap to obtain the confidence interval for hazard ratio. Defaults to TRUE.
n_boot	The number of bootstrap samples.
seed	The seed to reproduce the bootstrap results. The seed from the environment will be used if left unspecified.

Details

We use the following steps to obtain the hazard ratio estimate and confidence interval had there been no treatment switching:

- Use a pooled logistic regression switching model to estimate the causal parameter ψ based on the patients in the control group who had disease progression:

$$\text{logit}(p(E_{ik})) = \alpha U_{i,\psi} + \sum_j \beta_j x_{ijk}$$

where E_{ik} is observed switch status for individual i at observation k ,

$$U_{i,\psi} = T_{C_i} + e^{\psi} T_{E_i}$$

is the counterfactual survival time for individual i given a specific value for ψ , and x_{ijk} are all confounders for individual i at observation k . The visit-specific intercepts can be modeled using a natural cubic spline with specified degrees of freedom. The boundary knots and inner knots can be based on the range and percentiles of treatment switching times. When applied from a secondary baseline, $U_{i,\psi}$ refers to post-secondary baseline counterfactual survival, where T_{C_i} refers to the time spent after the secondary baseline on control treatment, and T_{E_i} refers to the time spent after the secondary baseline on the experimental treatment.

In the presence of censoring, let $U_{i,\psi} = T_{C_i} + e^{\psi} T_{E_i}$ and $V_{i,\psi} = \min(\tau_i, e^{\psi} \tau_i)$, where τ_i is the administrative censoring time for the subject. In addition, let Δ_i denote the observed event indicators, and let $W_{i,\psi} = \min(U_{i,\psi}, V_{i,\psi})$ and $\Delta_{i,\psi} = \Delta_i I(U_{i,\psi} \leq V_{i,\psi})$ be the recensored survival times and event indicators. Fit a null Cox model to $(W_{i,\psi}, \Delta_{i,\psi})$ to control patients with disease progression, and use the martingale residuals to replace the counterfactual survival times $U_{i,\psi}$ in the pooled logistic regression switching model.

- Search for ψ such that the estimate of α is close to zero. This will be the estimate of the causal parameter. The confidence interval for ψ can be obtained as the value of ψ such that the corresponding two-sided p-value for testing $H_0 : \alpha = 0$ in the switching model is equal to the nominal significance level.
- Derive the counter-factual survival times for control patients had there been no treatment switching. The counter-factual survival times are relative to randomization.
- Fit the Cox model to the observed survival times on the treatment arm and the counter-factual untreated survival times on the control arm to obtain the hazard ratio estimate.
- Use bootstrap to construct the p-value and confidence interval for hazard ratio.

Value

A list with the following components:

- `psi`: The estimated causal parameter for the control group.
- `psi_CI`: The confidence interval for `psi`.
- `psi_CI_type`: The type of confidence interval for `psi`, i.e., "logistic model" or "bootstrap".
- `logrank_pvalue`: The two-sided p-value of the log-rank test based on the treatment policy strategy.
- `cox_pvalue`: The two-sided p-value for treatment effect based on the Cox model.

- hr: The estimated hazard ratio from the Cox model.
- hr_CI: The confidence interval for hazard ratio.
- hr_CI_type: The type of confidence interval for hazard ratio, either "Cox model" or "bootstrap".
- settings: A list with the following components:
 - low_psi: The lower limit of the causal parameter.
 - hi_psi: The upper limit of the causal parameter.
 - strata_main_effect_only: Whether to only include the strata main effects in the logistic regression switching model.
 - firth: Whether the firth's penalized likelihood is used.
 - flic: Whether to apply intercept correction.
 - recensor: Whether to apply recensoring to counter-factual survival times.
 - admin_recensor_only: Whether to apply recensoring to administrative censoring time only.
 - swtrt_control_only: Whether treatment switching occurred only in the control group.
 - alpha: The significance level to calculate confidence intervals.
 - ties: The method for handling ties in the Cox model.
 - tol: The desired accuracy (convergence tolerance).
 - boot: Whether to use bootstrap to obtain the confidence interval for hazard ratio.
 - n_boot: The number of bootstrap samples.
 - seed: The seed to reproduce the simulation results.
- psi_trt: The estimated causal parameter for the treatment group if swtrt_control_only is FALSE.
- psi_trt_CI: The confidence interval for psi_trt if swtrt_control_only is FALSE.
- hr_boots: The bootstrap hazard ratio estimates if boot is TRUE.
- psi_boots: The bootstrap psi estimates if boot is TRUE.
- psi_trt_boots: The bootstrap psi_trt estimates if boot is TRUE and swtrt_control_only is FALSE.

Author(s)

Kaifeng Lu, <kaifenglu@gmail.com>

References

NR Latimer, IR White, K Tilling, and U Siebert. Improved two-stage estimation to adjust for treatment switching in randomised trials: g-estimation to address time-dependent confounding. *Statistical Methods in Medical Research*. 2020;29(10):2900-2918.

Examples

```
sim1 <- tsegestsim(
  n = 500, allocation1 = 2, allocation2 = 1, pbprog = 0.5,
  trtlghr = -0.5, bprogs1 = 0.3, shape1 = 1.8,
  scale1 = 0.000025, shape2 = 1.7, scale2 = 0.000015,
```

```

pmix = 0.5, admin = 5000, pcatnotrtbprog = 0.5,
pcattrtbprog = 0.25, pcatnotrt = 0.2, pcattrt = 0.1,
catmult = 0.5, tdxo = 1, ppoor = 0.1, pgood = 0.04,
ppoormet = 0.4, pgoodmet = 0.2, xomult = 1.4188308,
milestone = 546, swtrt_control_only = TRUE,
outputRawDataset = 1, seed = 2000)

fit1 <- tsegest(
  data = sim1$paneldata, id = "id",
  tstart = "tstart", tstop = "tstop", event = "died",
  treat = "trtrand", censor_time = "censor_time",
  pd = "progressed", pd_time = "timePFSobs", swtrt = "xo",
  swtrt_time = "xotime", swtrt_time_upper = "xotime_upper",
  base_cov = "bprog", conf_cov = "bprog*catlag",
  low_psi = -3, hi_psi = 3, strata_main_effect_only = TRUE,
  recensor = TRUE, admin_recensor_only = FALSE,
  swtrt_control_only = TRUE, alpha = 0.05, ties = "efron",
  tol = 1.0e-6, boot = FALSE)

c(fit1$hr, fit1$hr_CI)

sim2 <- tsegestsim(
  n = 500, allocation1 = 2, allocation2 = 1, pbprog = 0.5,
  trtlghr = -0.5, bprogs1 = 0.3, shape1 = 1.8,
  scale1 = 0.000025, shape2 = 1.7, scale2 = 0.000015,
  pmix = 0.5, admin = 5000, pcatnotrtbprog = 0.5,
  pcattrtbprog = 0.25, pcatnotrt = 0.2, pcattrt = 0.1,
  catmult = 0.5, tdxo = 1, ppoor = 0.1, pgood = 0.04,
  ppoormet = 0.4, pgoodmet = 0.2, xomult = 1.4188308,
  milestone = 546, swtrt_control_only = FALSE,
  outputRawDataset = 1, seed = 2000)

fit2 <- tsegest(
  data = sim2$paneldata, id = "id",
  tstart = "tstart", tstop = "tstop", event = "died",
  treat = "trtrand", censor_time = "censor_time",
  pd = "progressed", pd_time = "timePFSobs", swtrt = "xo",
  swtrt_time = "xotime", swtrt_time_upper = "xotime_upper",
  base_cov = "bprog", conf_cov = "bprog*catlag",
  low_psi = -3, hi_psi = 3, strata_main_effect_only = TRUE,
  recensor = TRUE, admin_recensor_only = FALSE,
  swtrt_control_only = FALSE, alpha = 0.05, ties = "efron",
  tol = 1.0e-6, boot = FALSE)

c(fit2$hr, fit2$hr_CI)

```

Description

Obtains the simulated data for baseline prognosis, disease progression, treatment switching, death, and time-dependent covariates.

Usage

```
tsegestsim(
  n = 500L,
  allocation1 = 2L,
  allocation2 = 1L,
  pbprog = 0.5,
  trtlghr = -0.5,
  bprogs1 = 0.3,
  shape1 = 1.8,
  scale1 = 2.5e-05,
  shape2 = 1.7,
  scale2 = 1.5e-05,
  pmix = 0.5,
  admin = 5000,
  pcatnotrtbprog = 0.5,
  pcatrtbprog = 0.25,
  pcatnotrt = 0.2,
  pcatrt = 0.1,
  catmult = 0.5,
  tdxo = 1,
  ppoor = 0.1,
  pgood = 0.04,
  ppoormet = 0.4,
  pgoodmet = 0.2,
  xomult = 1.4188308,
  milestone = 546,
  swtrt_control_only = 1L,
  outputRawDataset = 1L,
  seed = NA_integer_
)
```

Arguments

n	The total sample size for two treatment arms combined.
allocation1	The number of subjects in the active treatment group in a randomization block.
allocation2	The number of subjects in the control group in a randomization block.
pbprog	The probability of having poor prognosis at baseline.
trtlghr	The treatment effect in terms of log hazard ratio.
bprogs1	The poor prognosis effect in terms of log hazard ratio.
shape1	The shape parameter for the Weibull event distribution for the treatment group.
scale1	The scale parameter for the Weibull event distribution for the treatment group.

shape2	The shape parameter for the Weibull event distribution for the control group.
scale2	The scale parameter for the Weibull event distribution for the control group.
pmix	The mixing probability of the first component Weibull distribution.
admin	The administrative censoring time.
pcatnotrtbprog	The probability of developing metastatic disease on control treatment with poor baseline prognosis.
pcattrtbprog	The probability of developing metastatic disease on active treatment with poor baseline prognosis.
pcatnotrt	The probability of developing metastatic disease on control treatment with good baseline prognosis.
pcattrt	The probability of developing metastatic disease on active treatment with good baseline prognosis.
catmult	The impact of metastatic disease on shortening remaining survival time.
tdxo	whether treatment crossover occurs immediately after disease progression.
ppoor	The probability of switching for poor baseline prognosis with no metastatic disease.
pgood	The probability of switching for good baseline prognosis with no metastatic disease.
ppoormet	The probability of switching for poor baseline prognosis after developing metastatic disease.
pgoodmet	The probability of switching for good baseline prognosis after developing metastatic disease.
xomult	The direct effect of crossover on extending remaining survival time.
milestone	The milestone to calculate restricted mean survival time.
swtrt_control_only	Whether treatment switching occurred only in the control group.
outputRawDataset	whether to output the raw data set
seed	The seed to reproduce the simulation results. The seed from the environment will be used if left unspecified.

Value

A list with two data frames

- `sumdata`: A data frame with the following variables:
 - `simtrueconstmean`: The true control group restricted mean survival time (RMST).
 - `simtrueconstlb`: The lower bound for control group RMST.
 - `simtrueconstub`: The upper bound for control group RMST.
 - `simtrueconstse`: The standard error for control group RMST.
 - `simtrueexpstmean`: The true experimental group restricted mean survival time (RMST).
 - `simtrueexpstlb`: The lower bound for experimental group RMST.
 - `simtrueexpstub`: The upper bound for experimental group RMST.

- simtrueexpstse: The standard error for experimental group RMST.
- simtrue_coxwbprog_hr: The treatment hazard ratio from the Cox model adjusting for baseline prognosis.
- simtrue_cox_hr: The treatment hazard ratio from the Cox model without adjusting for baseline prognosis.
- paneldata: A counting process style data frame with the following variables:
 - id: The subject ID.
 - trtrand: The randomized treatment arm.
 - bprog: Whether the patient had poor baseline prognosis.
 - tstart: The left end of time interval.
 - tstop: The right end of time interval.
 - died: Whether the patient died.
 - progressed: Whether the patient had disease progression.
 - timePFSobs: The observed time of disease progression at regular scheduled visits.
 - progtdc: The time-dependent covariate for progression.
 - catevent: Whether the patient developed metastatic disease.
 - cattime: When the patient developed metastatic disease.
 - cattdc: The time-dependent covariate for cat event.
 - catlag: The lagged value of cattdc.
 - xo: Whether the patient switched treatment.
 - xotime: When the patient switched treatment.
 - xotdc: The time-dependent covariate for treatment switching.
 - xotime_upper: The upper bound of treatment switching time.
 - censor_time: The administrative censoring time.

Author(s)

Kaifeng Lu, <kaifenglu@gmail.com>

Examples

```
sim1 <- tsegestsim(
  n = 500, allocation1 = 2, allocation2 = 1, pbprog = 0.5,
  trtlghr = -0.5, bprogs1 = 0.3, shape1 = 1.8,
  scale1 = 0.000025, shape2 = 1.7, scale2 = 0.000015,
  pmix = 0.5, admin = 5000, pcatnotrtbprog = 0.5,
  pcatrtbprog = 0.25, pcatnotrt = 0.2, pcatrt = 0.1,
  catmult = 0.5, tdxo = 1, ppoor = 0.1, pgood = 0.04,
  ppoormet = 0.4, pgoodmet = 0.2, xomult = 1.4188308,
  milestone = 546, outputRawDataset = 1, seed = 2000)
```

`tsesimp`*The Simple Two-Stage Estimation (TSE) Method for Treatment Switching*

Description

Obtains the causal parameter estimate of the AFT model and the hazard ratio estimate of the Cox model to account for treatment switching.

Usage

```
tsesimp(  
  data,  
  stratum = "",  
  time = "time",  
  event = "event",  
  treat = "treat",  
  censor_time = "censor_time",  
  pd = "pd",  
  pd_time = "pd_time",  
  swtrt = "swtrt",  
  swtrt_time = "swtrt_time",  
  base_cov = "",  
  base2_cov = "",  
  aft_dist = "weibull",  
  strata_main_effect_only = TRUE,  
  recensor = TRUE,  
  admin_recensor_only = FALSE,  
  swtrt_control_only = TRUE,  
  alpha = 0.05,  
  ties = "efron",  
  offset = 1,  
  boot = TRUE,  
  n_boot = 1000,  
  seed = NA  
)
```

Arguments

`data` The input data frame that contains the following variables:

- `stratum`: The stratum.
- `time`: The survival time for right censored data.
- `event`: The event indicator, 1=event, 0=no event.
- `treat`: The randomized treatment indicator, 1=treatment, 0=control.
- `censor_time`: The administrative censoring time. It should be provided for all subjects including those who had events.

- pd: The disease progression indicator, 1=PD, 0=no PD.
- pd_time: The time from randomization to PD.
- swtrt: The treatment switch indicator, 1=switch, 0=no switch.
- swtrt_time: The time from randomization to treatment switch.
- base_cov: The values of baseline covariates (excluding treat).
- base2_cov: The values of baseline and secondary baseline covariates (excluding swtrt).

stratum	The name(s) of the stratum variable(s) in the input data.
time	The name of the time variable in the input data.
event	The name of the event variable in the input data.
treat	The name of the treatment variable in the input data.
censor_time	The name of the censor_time variable in the input data.
pd	The name of the pd variable in the input data.
pd_time	The name of the pd_time variable in the input data.
swtrt	The name of the swtrt variable in the input data.
swtrt_time	The name of the swtrt_time variable in the input data.
base_cov	The vector of names of base_cov variables (excluding treat) in the input data for the Cox model.
base2_cov	The vector of the names of base2_cov variables (excluding swtrt) in the input data for the AFT model.
aft_dist	The assumed distribution for time to event for the AFT model. Options include "exponential", "weibull", "loglogistic", and "lognormal".
strata_main_effect_only	Whether to only include the strata main effects in the AFT model. Defaults to TRUE, otherwise all possible strata combinations will be considered in the AFT model.
recensor	Whether to apply recensoring to counter-factual survival times. Defaults to TRUE.
admin_recensor_only	Whether to apply recensoring to administrative censoring time only. Defaults to FALSE, in which case, recensoring will be applied to the actual censoring time for dropouts.
swtrt_control_only	Whether treatment switching occurred only in the control group.
alpha	The significance level to calculate confidence intervals.
ties	The method for handling ties in the Cox model, either "breslow" or "efron" (default).
offset	The offset to calculate the time to event, PD, and treatment switch. We can set offset equal to 1 (default), 1/30.4375, or 1/365.25 if the time unit is day, month, or year.
boot	Whether to use bootstrap to obtain the confidence interval for hazard ratio. Defaults to TRUE.
n_boot	The number of bootstrap samples.
seed	The seed to reproduce the bootstrap results. The seed from the environment will be used if left unspecified.

Details

We use the following steps to obtain the hazard ratio estimate and confidence interval had there been no treatment switching:

- Use an AFT model to estimate the causal parameter ψ based on the patients in the control group who had disease progression.
- Derive the counter-factual survival times for control patients had there been no treatment switching.
- Fit the Cox model to the observed survival times on the treatment arm and the counter-factual untreated survival times on the control arm to obtain the hazard ratio estimate.
- Use bootstrap to construct the p-value and confidence interval for hazard ratio.

Value

A list with the following components:

- `psi`: The estimated causal parameter for the control group.
- `psi_CI`: The confidence interval for `psi`.
- `psi_CI_type`: The type of confidence interval for `psi`, i.e., "AFT model" or "bootstrap".
- `logrank_pvalue`: The two-sided p-value of the log-rank test based on the treatment policy strategy.
- `cox_pvalue`: The two-sided p-value for treatment effect based on the Cox model.
- `hr`: The estimated hazard ratio from the Cox model.
- `hr_CI`: The confidence interval for hazard ratio.
- `hr_CI_type`: The type of confidence interval for hazard ratio, either "model" or "bootstrap".
- `settings`: A list with the following components:
 - `aft_dist`: The distribution for time to event for the AFT model.
 - `strata_main_effect_only`: Whether to only include the strata main effects in the AFT model.
 - `recensor`: Whether to apply recensoring to counter-factual survival times.
 - `admin_recensor_only`: Whether to apply recensoring to administrative censoring time only.
 - `swtrt_control_only`: Whether treatment switching occurred only in the control group.
 - `alpha`: The significance level to calculate confidence intervals.
 - `ties`: The method for handling ties in the Cox model.
 - `offset`: The offset to calculate the time to event, PD, and treatment switch.
 - `boot`: Whether to use bootstrap to obtain the confidence interval for hazard ratio.
 - `n_boot`: The number of bootstrap samples.
 - `seed`: The seed to reproduce the simulation results.
- `psi_trt`: The estimated causal parameter for the treatment group if `swtrt_control_only` is FALSE.
- `psi_trt_CI`: The confidence interval for `psi_trt` if `swtrt_control_only` is FALSE.
- `hr_boots`: The bootstrap hazard ratio estimates if `boot` is TRUE.

- `psi_boots`: The bootstrap `psi` estimates if `boot` is `TRUE`.
- `psi_trt_boots`: The bootstrap `psi_trt` estimates if `boot` is `TRUE` and `swtrt_control_only` is `FALSE`.

Author(s)

Kaifeng Lu, <kaifenglu@gmail.com>

References

Nicholas R Latimer, KR Abrams, PC Lambert, MK Crowther, AJ Wailoo, JP Morden, RL Akehurst, and MJ Campbell. Adjusting for treatment switching in randomised controlled trials - A simulation study and a simplified two-stage method. *Statistical Methods in Medical Research*. 2017;26(2):724-751.

Examples

```
library(dplyr)

# the eventual survival time
shilong1 <- shilong %>%
  arrange(bras.f, id, tstop) %>%
  group_by(bras.f, id) %>%
  filter(row_number() == n()) %>%
  select(-c("ps", "ttc", "tran"))

# the last value of time-dependent covariates before pd
shilong2 <- shilong %>%
  filter(pd == 0 | tstart <= dpd) %>%
  arrange(bras.f, id, tstop) %>%
  group_by(bras.f, id) %>%
  filter(row_number() == n()) %>%
  select(bras.f, id, ps, ttc, tran)

# combine baseline and time-dependent covariates
shilong3 <- shilong1 %>%
  left_join(shilong2, by = c("bras.f", "id"))

# apply the two-stage method
fit <- tsesimp(
  data = shilong3, time = "tstop", event = "event",
  treat = "bras.f", censor_time = "dcut", pd = "pd",
  pd_time = "dpd", swtrt = "co", swtrt_time = "dco",
  base_cov = "",
  base2_cov = c("agerand", "sex.f", "tt_Lnum", "rmh_alea.c",
                "pathway.f", "ps", "ttc", "tran"),
  aft_dist = "weibull", alpha = 0.05,
  recensor = TRUE, swtrt_control_only = FALSE, offset = 1,
  boot = FALSE)

c(fit$hr, fit$hr_CI)
```

Index

* datasets

- aml, [3](#)
 - heart, [5](#)
 - immdef, [5](#)
 - ingots, [6](#)
 - rawdata, [32](#)
 - sexagg, [41](#)
 - shilong, [42](#)
 - six, [43](#)
 - tobin, [47](#)
-
- aml, [3](#)
-
- bscpp, [4](#)
-
- heart, [5](#)
-
- immdef, [5](#)
 - ingots, [6](#)
 - ipcw, [7](#)
 - ipe, [11](#)
-
- kmest, [15](#)
-
- liferegr, [17](#)
 - logisregr, [21](#)
 - lrtest, [24](#)
-
- nscpp, [26](#)
-
- phregr, [27](#)
-
- qrcpp, [31](#)
-
- rawdata, [32](#)
 - residuals_phregr, [33](#)
 - rmdiff, [34](#)
 - rmest, [36](#)
 - rpsftm, [37](#)
-
- sexagg, [41](#)
 - shilong, [42](#)
-
- six, [43](#)
 - splineDesigncpp, [44](#)
 - survfit_phregr, [45](#)
-
- tobin, [47](#)
 - trtswitch-package, [2](#)
 - tsegest, [47](#)
 - tsegestsim, [52](#)
 - tsesimp, [56](#)