Package 'SimHaz'

August 31, 2015

Version 0.1

Date 2015-07-22

Title Simulated Survival and Hazard Analysis for time-dependent exposure

Author Danyi Xiong, Teeranan Pokaprakarn, Nusrat Rabbee

Maintainer Nusrat Rabbee <rabbee@berkeley.edu>

Depends R (>= 3.1.1), survival

Suggests MASS

Description SimHaz package in R to generate power for the Cox proportional hazards model by simulating survival events data with time dependent exposure status for subjects. A dichotomous exposure variable is considered with a single transition from unexposed to exposed status during the subject's time on study.

License GPL (>= 2)

URL http://www.r-project.org, http://www.another.url

BugReports http://pkgname.bugtracker.url

R topics documented:

SimHaz-package *Simulated Survival and Hazard Analysis for time-dependent*

Description

SimHaz package in R to generate power for the Cox proportional hazards model by simulating survival events data with time dependent exposure status for subjects. A dichotomous exposure variable is considered with a single transition from unexposed to exposed status during the subject time's on study.

Details

Package:	SimHaz
Type:	Package
Version:	0.1
Date:	2015-07-22
License:	GPL-2
Depends: R (>= 3.1.1), survival	

Author(s)

Danyi Xiong, Teeranan Pokaprakarn, Nusrat Rabbee
Maintainer: Nusrat Rabbee <rabbee@berkeley.edu>

References

~~ Literature or other references for background information ~~

See Also

~~ Optional links to other man pages, e.g. ~~ ~~ <pkg> ~~

Examples

```
~~ simple examples of the most important functions ~~
```

getpower.clst *Calculate power for the Cox proportional hazard model with time-dependent exposure using method 1 with clustering*

Description

This functions runs nSim (number of simulations, specified by the user) Monte Carlo simulations on the Cox proportional model with the cluster option. At each simulation, the function calls tdSim.clst internally. The function returns a data frame of scenario-specific parameters (including statistical power) and appends the output to a file with file name specified in the input parameters list. The user also have an option whether to plot an incidence plot or not.

Usage

```
getpower.clst(nSim, N, duration = 24, med.TTE.Control = 24, rho = 1, beta, med.TimeToCensor = 14, d
```

Arguments

nSim	Number of simulations
N	Number of subjects needs to be screened
duration	Length of the study in Months. The default value is 24 (months)
med.TTE.Control	Median time to event for control group. The default value is 24 (months)
rho	Shape parameter of the Weibull distribution, which is defaulted as 1, as we generate survival times by using the exponential distribution.
beta	A numeric value that represents the exposure effect, which is the regression coefficient (log hazard ratio) that represent the magnitude of the relationship between the exposure covariate and the risk of an event.
med.TimeToCensor	Median time to censoring for all subjects. The default value is 14 (months). See also at help document for tdSim.method1.
df	A user-specified n (n 3) by 3 clustering data frame with columns corresponding to cat_id (category id, which is the physician site id. It can be either text strings or integers), cat_prop (category proportion, which is the proportion of subjects in corresponding a category id), and cat_exprate (category exposure rate, which is the exposure proportion corresponding to a category id). n rows corresponds to n different physician sites.
type	A text string indicating the what type of dataset is of interest. Either one of <e2><80><9c>fixed<e2><80><9d> or <e2><80><9c>td<e2><80><9d> should be inputted
scenario	Any text string inputted by the user as an option to name a scenario that is being simulated. The use can simply put <e2><80><9c> <e2><80><9d> if he/she decides to not name the scenario
prop.fullexp	A numeric value in interval [0, 1) that represents the proportion of exposed subjects that are fully exposed from the beginning to the end of the study. The default value is 0, which means all exposed subjects have an exposure status transition at some point during the study
maxrelexptime	A numeric value in interval (0, 1] that represents the maximum relative exposure time. The default value is 1, which means all exposed subjects have an exposure status transition at some point during the study and possibly at the start or end
min.futime	A numeric value that represents minimum follow-up time (in months). The default value is 0, which means no minimum follow-up time is considered. If it has a positive value, this argument will help exclude subjects that only spend a short amount of time in the study
min.postexp.futime	A numeric value that represents minimum post-exposure follow-up time (in months). The default value is 0, which means no minimum post-exposure follow-up time is considered. If it has a positive value, this argument will help exclude subjects that only spend a short amount of time in the study after their exposure
output.fn	A .csv filename to write in the output. If the filename does not exist, the function will create a new .csv file for the output
incidence.plot	A logical value indicating whether to output an incidence plot or not.The default value is FALSE

Details

The user needs to install the survival package in order to use this function, as it calculates power based on the Cox regression model, which calls the coxph function from the survival library.

Value

A data.frame object with 3 rows and columns corresponding to

i_scenario	Scenario name specified by the user
i_type	Dataset type specified by the user
i_N	Number of subjects needs to be screened, specified by the user
i_min.futime	Minimum follow-up time to be considered, specified by the user
i_min.postexp.futime	Minimum post-exposure follow-up time to be considered, specified by the user
i_cat	Category id specified in user<e2><80><99>s input data frame
i_cat_prop	Category proportion specified in user<e2><80><99>s input data frame
i_cat_exp.prop	Category exposure proportion specified in user<e2><80><99>s input data frame
i_exp.prop	Weighted exposure proportion calculated from user<e2><80><99>s input data frame
i_lambda	Value of the scale parameter of the Weibull distribution to generate survival times. Calculated from median time to event for control group, which is specified by the user
i_rho	User-specified Value of the shape parameter of the Weibull distribution to generate survival times
i_rateC	Rate of the exponential distribution to generate censoring times. Calculated from median time to censoring, which is specified by the user. i_beta Input value of regression coefficient (log hazard ratio)
N_eff	Simulated number of evaluable subjects, which is the resulting number of subjects with or without considering minimum follow-up time and/or minimum post-exposure follow-up time
N_effexp_p	Simulated proportion of exposed subjects with or without considering minimum follow-up time and/or minimum post-exposure follow-up time
bhat	Simulated value of regression coefficient (log hazard ratio)
HR	Simulated value of hazard ratio
d	Simulated number of events in total
d_c	Simulated number of events in control group
d_exp	Simulated number of events in exposed group
mst_c	Simulated median survival time in control group
mst_exp	Simulated median survival time in exposed group
pow	Simulated statistical power from the Cox regression model on data with time-dependent exposure

References

Savignoni et al.: Matching methods to create paired survival data based on an exposure occurring over time: a simulation study with application to breast cancer. BMC Medical Research Methodology 2014 14:83.

Examples

```
# Install the survival package if needed.

library(survival)

# Create a clustering data frame as input with 3 categories and a 20% weighted
# exposure proportion.

input_df <- data.frame(cat_id = c('lo', 'med', 'hi'), cat_prop = c(0.65, 0.2, 0.15),
cat_exp.prop = c(0.1, 0.3, 0.5))

# Run 500 simulations. Each time simulate a dataset of 600 subjects with
# time-dependent exposure with both minimum follow-up time (4 months) and
# minimum post-exposure follow-up time (4 months) imposed. Also consider a
# quick exposure after entering the study for each exposed subject. Set the
# maximum relative exposure time to be <e2><85><99>.

# Set the duration of the study to be 24 months; the median time to event for
# control group to be 24 months; exposure effect to be 0.3; median time to
# censoring to be 14 months.

ret <- getpower.clst(nSim = 500, N = 600, beta = 0.3, df = input_df, type = "td", scenario = "clustering", maxr
```

`getpower.method1` *Calculate power for the Cox proportional hazard model with time-dependent exposure using method 1*

Description

This functions runs nSim (Number of simulations, specified by the user) Monte Carlo simulations, each time it calls tdSim.method1 internally. The function returns a data frame of scenario-specific parameters (including statistical power) and appends the output to a file with file name specified in the input parameters list. The user also have an option whether to plot an incidence plot or not.

Usage

```
getpower.method1(nSim, N, duration = 24, med.TTE.Control = 24, rho = 1, med.TimeToCensor = 14,
b, exp.prop, type, scenario, prop.fullexp = 0, maxrelexptime = 1, min.futime = 0,
min.postexp.futime = 0, output.fn, incidence.plot = FALSE)
```

Arguments

`nSim`	Number of simulations
`N`	Number of subjects needs to be screened
`duration`	Length of the study in Months. The default value is 24 (months)
`med.TTE.Control`	Median time to event for control group. The default value is 24 (months)
`rho`	Shape parameter of the Weibull distribution, which is defaulted as 1, as we generate survival times by using the exponential distribution. med.TimeToCensor Median time to censoring for all subjects. The default value is 14 (months) \itmemed.TimeToCensorMedian time to censoring for all subjects. The default value is 14 (months)

b	A numeric value that represents the exposure effect, which is the regression coefficient (log hazard ratio) that represent the magnitude of the relationship between the exposure covariate and the risk of an event
exp.prop	A numeric value between 0 and 1 (not include 0 and 1) that represents the proportion of subjects that are assigned with an exposure
type	A text string indicating the what type of dataset is of interest. Either one of <e2><80><9c>fixed<e2><80><9d> or <e2><80><9c>td<e2><80><9d> should be inputted
scenario	Any text string inputted by the user as an option to name a scenario that is being simulated. The use can simply put <e2><80><9c> <e2><80><9d> if he/she decides to not name the scenario
prop.fullexp	A numeric value in interval [0, 1) that represents the proportion of exposed subjects that are fully exposed from the beginning to the end of the study. The default value is 0, which means all exposed subjects have an exposure status transition at some point during the study
maxrelexptime	A numeric value in interval (0, 1] that represents the maximum relative exposure time. The default value is 1, which means all exposed subjects have an exposure status transition at some point during the study and possibly at the start or end
min.futime	A numeric value that represents minimum follow-up time (in months). The default value is 0, which means no minimum follow-up time is considered. If it has a positive value, this argument will help exclude subjects that only spend a short amount of time in the study
min.postexp.futime	A numeric value that represents minimum post-exposure follow-up time (in months). The default value is 0, which means no minimum post-exposure follow-up time is considered. If it has a positive value, this argument will help exclude subjects that only spend a short amount of time in the study after their exposure
output.fn	A .csv filename to write in the output. If the filename does not exist, the function will create a new .csv file for the output
incidence.plot	A logical value indicating whether to output an incidence plot or not.The default value is FALSE

Details

The user needs to install the survival package in order to use this function, as it calculates power based on the Cox regression model, which calls the coxph function from the survival library

Value

A data.frame object with columns corresponding to

i_scenario	Scenario name specified by the user
i_type	Dataset type specified by the user
i_N	Number of subjects needs to be screened, specified by the user
i_min.futime	Minimum follow-up time to be considered, specified by the user
i_min.postexp.futime	Minimum post-exposure follow-up time to be considered, specified by the user
i_exp.prop	Exposure rate specified by the user

i_lambda	Value of the scale parameter of the Weibull distribution to generate survival times. Calculated from median time to event for control group, which is specified by the user
i_rho	User-specified Value of the shape parameter of the Weibull distribution to generate survival times
i_rateC	Rate of the exponential distribution to generate censoring times. Calculated from median time to censoring, which is specified by the user. i_beta Input value of regression coefficient (log hazard ratio)
N_eff	Simulated number of evaluable subjects, which is the resulting number of subjects with or without considering minimum follow-up time and/or minimum post-exposure follow-up time
N_effexp_p	Simulated proportion of exposed subjects with or without considering minimum follow-up time and/or minimum post-exposure follow-up time
bhat	Simulated value of regression coefficient (log hazard ratio)
HR	Simulated value of hazard ratio
d	Simulated number of events in total
d_c	Simulated number of events in control group
d_exp	Simulated number of events in exposed group
mst_c	Simulated median survival time in control group
mst_exp	Simulated median survival time in exposed group
pow	Simulated statistical power from the Cox regression model on data with time-dependent exposure

Author(s)

Danyi Xiong, Teeranan Pokaprakarn, Nusrat Rabbee
Maintainer: Nusrat Rabbee <rabbee@berkeley.edu>

References

Therneau T (2015). A Package for Survival Analysis in S. version 2.38, http://CRAN.R-project.org/package=survival

Examples

```
# Install the survival package if needed.

library(survival)

# Run 500 simulations. Each time simulate a dataset of 600 subjects with
# time-dependent exposure with both minimum follow-up time (4 months) and
# minimum post-exposure follow-up time (4 months) imposed. Also consider a
# quick exposure after entering the study for each exposed subject. Set the
# maximum relative exposure time to be <e2><85><99>.

# Set the duration of the study to be 24 months; the median time to event for
# control group to be 24 months; exposure effect to be 0.3; median time to
# censoring to be 14 months; and exposure proportion to be 20%.

ret <- getpower.method1(nSim = 500, N = 600, b = 0.3, exp.prop = 0.2, type = "td", scenario =
" ", maxrelexptime = 1/6, min.futime = 4, min.postexp.futime = 4,
output.fn = "output.csv")
```

`getpower.method2` *Calculate power for the Cox proportional hazard model with time-dependent exposure using method 1*

Description

This functions runs nSim (Number of simulations, specified by the user) Monte Carlo simulations, each time it calls tdSim.method2 internally. The function returns a data frame of scenario-specific input parameters- and it also output statistical power. The user has the option to append the output to a file with file name specified in the input parameters list.

Usage

```
getpower.method2(nSim = 500, N, duration = 24, scenario, lambda12, lambda23 = NULL, lambda13, HR =
```

Arguments

`nSim`	Number of simulations
`N`	Number of subjects needs to be screened
`duration`	Length of the study in Months. The default value is 24 (months)
`scenario`	Any text string inputted by the user as an option to name a scenario that is being simulated. The use can simply put <e2><80><9c> <e2><80><9d> if he/she decides to not name the scenario
`lambda12`	Lambda12 parameter to control time to exposure
`lambda23`	Lambda23 parameter to control time to event after exposure
`lambda13`	Lambda13 parameter to control time to event in the control group
`HR`	Hazard Ratio. This input is optional. If HR is set and lambda23 is not set, lambda23 = lambda13*HR
`exp.prop`	A numeric value between 0 and 1 (not include 0 and 1) that represents the proportion of subjects that are assigned with an exposure
`rateC`	Rate of the exponential distribution to generate censoring times
`min.fut`	A numeric value that represents minimum follow-up time (in months). The default value is 0, which means no minimum follow-up time is considered. If it has a positive value, this argument will help exclude subjects that only spend a short amount of time in the study
`min.postexp.fut`	A numeric value that represents minimum post-exposure follow-up time (in months). The default value is 0, which means no minimum post-exposure follow-up time is considered. If it has a positive value, this argument will help exclude subjects that only spend a short amount of time in the study after their exposure
	A .csv filename to write in the output. If the filename does not exist, the function will create a new .csv file for the output

Examples

```
##---- Should be DIRECTLY executable !! ----
##-- ==>  Define data, use random,
##--or do  help(data=index)  for the standard data sets.

## The function is currently defined as
function (nSim = 500, N, duration = 24, scenario, lambda12, lambda23 = NULL,
    lambda13, HR = NULL, exp.prop, rateC, min.fut, min.postexp.fut,
    filename, simu.plot = FALSE)
{
    set.seed(999)
    try(if (lambda23 == NULL & HR == NULL) {
        stop("either lambda23 or HR(Hazard ratio) must be set")
    })
    if (lambda23 == NULL & HR != NULL) {
        lambda23 = lambda13 * HR
    }
    res = matrix(0, numsim, 10)
    colnames(res) = c("N.eff", "N.effexp.p", "betahat", "HR",
        "signif", "events", "events_c", "events_exp", "medsurvt_c",
        "medsurvt_exp")
    alpha = 0.05
    if (simu.plot) {
        dat <- tdSim.method2(N, duration, lambda12 = lambda12,
            lambda23 = lambda23, lambda13 = lambda, exp.prop = exp.prop,
            rateC = rateC, min.fut = min.fut, min.postexp.fut = min.postexp.fut)
        plot_simuData(dat)
    }
    for (k in 1:nSim) {
        dat <- tdSim.method2(N, duration, lambda12 = lambda12,
            lambda23 = lambda23, lambda13 = lambda, exp.rate = exp.prop,
            rateC = rateC, min.fut = min.fut, min.postexp.fut = min.postexp.fut)
        fit <- coxph(Surv(start, stop, status) ~ factor(x), data = dat)
        sfit <- survfit(Surv(start, stop, status) ~ factor(x),
            data = dat)
        res[k, "N.eff"] <- length(unique(dat$id))
        res[k, "N.effexp.p"] <- sum(dat$x)/length(unique(dat$id))
        res[k, "betahat"] <- summary(fit)$coef[, "coef"]
        res[k, "HR"] <- summary(fit)$coef[, "exp(coef)"]
        res[k, "signif"] <- ifelse(summary(fit)$coef[, "Pr(>|z|)"] <
            alpha, 1, 0)
        res[k, "events"] <- sum(dat$status)
        res[k, "events_c"] <- summary(sfit)$table[1, "events"]
        res[k, "events_exp"] <- summary(sfit)$table[2, "events"]
        res[k, "medsurvt_c"] <- summary(sfit)$table[1, "median"]
        res[k, "medsurvt_exp"] <- summary(sfit)$table[2, "median"]
    }
    df = data.frame(i_scenario = scenario, i_N = N, i_min.postexp.fut = min.postexp.fut,
        i_min.postexp.fut = min.postexp.fut, i_exprate = er,
        i_lambda12 = lambda12, i_lambda23 = lambda23, i_lambda13 = lambda13,
        i_rateC = rateC, N_eff = mean(res[, "N.eff"]), N_effexp_p = mean(res[,
            "N.effexp.p"]), bhat = mean(res[, "betahat"]), HR = mean(res[,
            "HR"]), d = mean(res[, "events"]), d_c = mean(res[,
            "events_c"]), d_exp = mean(res[, "events_exp"]),
        mst_c = mean(na.omit(res[, "medsurvt_c"])), mst_exp = mean(na.omit(res[,
            "medsurvt_exp"])), pow = mean(res[, "signif"]))
```

```
  if (file.exists(filename)) {
      write.table(df, file = filename, row.names = FALSE, col.names = FALSE,
          append = TRUE, sep = ",")
  }
  else {
      write.table(df, file = filename, row.names = FALSE, col.names = TRUE,
          sep = ",")
  }
  return(df)
}
```

`plot.incidence` *Make an incidence plot from simulated data.*

Description

Create an incidence plot ordered by follow-up time from a survival data simulated.

Usage

```
plot.incidence(data, title="Sample Survival Data")
```

Arguments

`data`	A dataframe of survival data containing the following columns: id, start, stop, status, x
`title`	Title of the graph

Details

This makes the incidence plot of the survival data based on the input dataframe from the tdSim.method1 or tdSim.method2 function. More generally, this function would also works with a dataframe containing survival data with the same columns name as indicated above.

Author(s)

Danyi Xiong, Teeranan Pokaprakarn, Nusrat Rabbee
Maintainer: Nusrat Rabbee <rabbee@berkeley.edu>

Examples

```
dat <- tdSim.method2(500, lambda12=1.36, lambda23=0.0389, lambda13=0.0288, exp.rate=0.2,rateC=0.05, filterA=
plot_simuData(dat, title='method2_filter')
```

plot_power *Plot power curves for survival analysis with time-dependent exposure*

Description

This function plots a power curve at each time and returns a subsetted data frame that match the list of input parameters.

Usage

```
plot_power(table_df, N, type, exp.prop, min.futime, min.postexp.futime, plot = TRUE,
newplot, col, lty, lwd, pch)
```

Arguments

table_df	A data frame read from a .csv file of a format output from the getpower.method1 function.
N	Number of subjects needs to be screened
type	A text string indicating the what type of dataset is of interest. Either one of <e2><80><9c>fixed<e2><80><9d> or <e2><80><9c>td<e2><80><9d> should be inputted
exp.prop	A numeric value between 0 and 1 (not include 0 and 1) that represents the proportion of subjects that are assigned with an exposure
min.futime	A numeric value that represents minimum follow-up time (in months). The default value is 0, which means no minimum follow-up time is considered. If it has a positive value, this argument will help exclude subjects that only spend a short amount of time in the study
min.postexp.futime	A numeric value that represents minimum post-exposure follow-up time (in months). The default value is 0, which means no minimum post-exposure follow-up time is considered. If it has a positive value, this argument will help exclude subjects that only spend a short amount of time in the study after their exposure
plot	A logical value indicating whether to output a power curve or not. The default value is TRUE
newplot	A logical value indicating whether to create a new plot or add to an existing plot
col,lty,lwd,pch	Graphical parameters as in the regular plot function in R

Details

The gist of this function is that the user can check the plot with the values in the output data frame. Moreover, it is flexible that the user can choose to only output the data frame in order to plot their own graph (such as add titles or legends) based on the data. The user can also add as many lines as possible to an existing power curve plot so that he/she is allowed to compare different scenarios.

Value

A data.frame object with columns corresponding to

i_N	Number of subjects needs to be screened, specified by the user
N_eff	Simulated number of evaluable subjects, which is the resulting number of subjects with or without considering minimum follow-up time and/or minimum post-exposure follow-up time
i_beta	Input value of regression coefficient (log hazard ratio)
pow	Simulated statistical power from the Cox regression model on data with time-dependent exposure

Examples

```
# Read in .csv file as a data frame

  tb <-  read.csv("output.csv", header = TRUE, sep = ",")

# Visualize the subsetted data frame of interest and create a new plot

visualize1 <- plot_power(table_df = tb, N = 600, type = "td", exp.prop = 0.2,
min.futime = 4, min.postexp.futime = 4, plot = TRUE, newplot = TRUE, co l = "red", lty = 1, lwd = 2, pch = 16)

# Add a different power curve to the previously created plot

visualize2 <- plot_power(table_df = tb, N = 600, type = "td", exp.prop=0.2,
min.futime = 4, min.postexp.futime = 0, plot = TRUE, newplot = FALSE, col = "blue", lty = 1, lwd = 2, pch = 16)
```

---

tdSim.clst *Simulate 1 dataframe (1 simulation) of time-dependent exposure under method 1 with a clustering data frame*

Description

This function allows the user to input a data frame with clustering parameters and generates a simulated dataset with time-dependent exposure. In particular, the output dataset has a column corresponding to the physician site id, which will be used as a clustering variable in the Cox regression model in power calculation.

Usage

```
tdSim.clst(N, duration = 24, lambda, rho = 1, beta, rateC, df, prop.fullexp = 0, maxrelexptime = 1,
```

Arguments

N	Number of subjects needs to be screened
duration	Length of the study in Months. The default value is 24 (months)
lambda	Scale parameter of the Weibull distribution, which is calculated as log(2) / median time to event for control group
rho	Shape parameter of the Weibull distribution, which is defaulted as 1, as we generate survival times by using the exponential distribution

beta	A numeric value that represents the exposure effect, which is the regression coefficient (log hazard ratio) that represent the magnitude of the relationship between the exposure covariate and the risk of an event
rateC	Rate of the exponential distribution to generate censoring times, which is calculated as log(2) / median time to censoring
df	A user-specified n (n 3) by 3 clustering data frame with columns corresponding to cat_id (category id, which is the physician site id. It can be either text strings or integers), cat_prop (category proportion, which is the proportion of subjects in corresponding a category id), and cat_exprate (category exposure rate, which is the exposure proportion corresponding to a category id). n rows corresponds to n different physician sites
prop.fullexp	A numeric value in interval [0, 1) that represents the proportion of exposed subjects that are fully exposed from the beginning to the end of the study. The default value is 0, which means all exposed subjects have an exposure status transition at some point during the study
maxrelexptime	A numeric value in interval (0, 1] that represents the maximum exposure time relative to the follow up time. The default value is 1, which means that exposure time is uniformly distributed between start and follow-up time.
min.futime	A numeric value that represents minimum follow-up time (in months). The default value is 0, which means no minimum follow-up time is considered. If it has a positive value, this argument will help exclude subjects that only spend a short amount of time in the study
min.postexp.futime	A numeric value that represents minimum post-exposure follow-up time (in months). The default value is 0, which means no minimum post-exposure follow-up time is considered. If it has a positive value, this argument will help exclude subjects that only spend a short amount of time in the study after their exposure

Details

The current version of this function allows the user to input a data frame with at least 3 categories of physician sites, because the function uses a multinomial distribution to assign subjects into each category according to the corresponding category proportion

Value

A data.frame object with columns corresponding to

id	Integer that represents a subject<e2><80><99>s identification number
start	For counting process formulation. Represents the start of each time interval
stop	For counting process formulation. Represents the end of each time interval
status	Indicator of event. status = 1 when event occurs and 0 otherwise
x	Indicator of exposure. x = 1 when exposed and 0 otherwise
clst_id	For clustering in the Cox proportional hazard model. Represents label of each subject<e2><80><99>s corresponding physician site

Author(s)

Danyi Xiong, Teeranan Pokaprakarn, Nusrat Rabbee
Maintainer: Nusrat Rabbee <rabbee@berkeley.edu>

References

T. Therneau and C. Crowson (2015). Using Time Dependent Covariates and Time Dependent Coefficients in the Cox Model. https://cran.r-project.org/web/packages/survival/vignettes/timedep.pdf

Examples

```
# Create a clustering data frame as input with 3 categories and a 20% weighted
# exposure proportion.

input_df <- data.frame(cat_id = c('lo', 'med', 'hi'), cat_prop = c(0.65, 0.2, 0.15),
cat_exp.prop = c(0.1, 0.3, 0.5))

# Simulate a dataset of 600 subjects with time-dependent exposure. Consider
# both minimum follow-up time (4 months) and minimum post-exposure follow-up # time (4 months). Also consider
# each exposed subject. Set the maximum relative exposure time to be <e2><85><99>.

# Set the duration of the study to be 24 months; the median time to event for
# control group to be 24 months; exposure effect to be 0.3; median time to
# censoring to be 14 months.

df_tdclst <- tdSim.clust(N = 600, duration = 24, lambda = log(2)/24, rho = 1,
beta = 0.3, rateC = log(2)/14, df = input_df, prop.fullexp = 0, maxrelexptime = 1/6,
min.futime = 4, min.postexp.futime = 4)
```

`tdSim.method1` *Simulate 1 dataframe (1 simulation) of time-dep exposure under method 1*

Description

This function generates a simulated dataset with time-dependent exposure under method 1 with a user-specified list of parameters as input. Survival times and censoring times are generated from the exponential distribution.

Usage

```
tdSim.method1(N, duration = 24, lambda, rho = 1, beta, rateC, exp.prop,
prop.fullexp  = 0, maxrelexptime = 1, min.futime = 0, min.postexp.futime = 0)
```

Arguments

`N`	Number of subjects needs to be screened
`duration`	Length of the study in Months. The default value is 24 (months)
`lambda`	Scale parameter of the Weibull distribution, which is calculated as log(2) / median time to event for control group
`rho`	Shape parameter of the Weibull distribution, which is defaulted as 1, as we generate survival times by using the exponential distribution
`beta`	A numeric value that represents the exposure effect, which is the regression coefficient (log hazard ratio) that represent the magnitude of the relationship between the exposure covariate and the risk of an event

`rateC` Rate of the exponential distribution to generate censoring times, which is calculated as log(2) / median time to censoring

`exp.prop` A numeric value between 0 and 1 (not include 0 and 1) that represents the proportion of subjects that are assigned with an exposure

`prop.fullexp` A numeric value in interval [0, 1) that represents the proportion of exposed subjects that are fully exposed from the beginning to the end of the study. The default value is 0, which means all exposed subjects have an exposure status transition at some point during the study

`maxrelexptime` A numeric value in interval (0, 1] that represents the maximum exposure time relative to the follow up time. The default value is 1, which means that exposure time is uniformly distributed between start and follow-up time.

`min.futime` A numeric value that represents minimum follow-up time (in months). The default value is 0, which means no minimum follow-up time is considered. If it has a positive value, this argument will help exclude subjects that only spend a short amount of time in the study

`min.postexp.futime` A numeric value that represents minimum post-exposure follow-up time (in months). The default value is 0, which means no minimum post-exposure follow-up time is considered. If it has a positive value, this argument will help exclude subjects that only spend a short amount of time in the study after their exposure

Details

If no minimum follow-up time or minimum post-exposure follow-up time is considered (min.fut = 0 and min.postexp.fut = 0), then the output data frame will have N subjects. If we consider minimum follow-up time or both, then the output data frame will have at most N subjects.

Value

A data.frame object with columns corresponding to

`id` Integer that represents a subject<e2><80><99>s identification number

`start` For counting process formulation. Represents the start of each time interval

`stop` For counting process formulation. Represents the end of each time interval

`status` Indicator of event. status = 1 when event occurs and 0 otherwise

`x` Indicator of exposure. x = 1 when exposed and 0 otherwise

Author(s)

Danyi Xiong, Teeranan Pokaprakarn, Nusrat Rabbee
Maintainer: Nusrat Rabbee <rabbee@berkeley.edu>

References

Therneau and C. Crowson (2015). Using Time Dependent Covariates and Time Dependent Coefficients in the Cox Model. https://cran.r-project.org/web/packages/survival/vignettes/timedep.pdf

Examples

```
# Simulate a dataset of 600 subjects with time-dependent exposure without
# considering minimum follow-up time or minimum post-exposure follow-up time.
# Specifically, set the duration of the study to be 24 months; the median time to
# event for control group to be 24 months; exposure effect to be 0.3; median time
# to censoring to be 14 months; and exposure proportion to be 20%.

df1 <- tdSim.method1(N = 600, duration = 24, lambda = log(2)/24, rho = 1,
   beta = 0.3, rateC = log(2)/14, exprate = 0.2,
   fullyexp.p  = 0, maxrelexp.t = 1, min.fut = 0,
   min.postexp.fut = 0)

# Simulate a dataset of 600 subjects with time-dependent exposure with
# both minimum follow-up time (4 months) and minimum post-exposure
# follow-up time (4 months) imposed. Other parameters remain the same as
# in the first case.

df2 <- tdSim.method1(N = 600, duration = 24, lambda = log(2)/24, rho = 1,
   beta = 0.3, rateC = log(2)/14, exprate = 0.2,
   fullyexp.p  = 0, maxrelexp.t = 1, min.fut = 4,
   min.postexp.fut = 4)

# Simulate a dataset of 600 subjects with time-dependent exposure with
# both minimum follow-up time (4 months) and minimum post-exposure
# follow-up time (4 months) imposed. Also consider a quick exposure after entering the study for each exposed

df3 <- tdSim.method1(N = 600, duration = 24, lambda = log(2)/24, rho = 1,
   beta = 0.3, rateC = log(2)/14, exprate = 0.2,
   fullyexp.p  = 0, maxrelexp.t = 1/6, min.fut = 4,
   min.postexp.fut = 4)
```

`tdSim.method2` *Simulate 1 dataframe (1 simulation) of time-dep exposure under method 2*

Description

This function simulate Survival Data. It generates a simulated dataset with time-dependent exposure under method 2

Usage

```
tdSim.method2 <-function(N,duration, lambda12, lambda23=NULL, lambda13,
HR=NULL, exp.prop,rateC, min.futime = 0, min.postexp.futime = 0)
```

Arguments

`N`	Number of subjects
`duration`	Duration of the study. This is used in censoring
`lambda12`	Lambda12 parameter to control time to exposure
`lambda23`	Lambda23 parameter to control time to event after exposure
`lambda13`	Lambda13 parameter to control time to event in the control group

HR	Hazard Ratio. This input is optional. If HR is set and lambda23 is not set, lambda23 = lambda13*HR
exp.prop	A numeric value between 0 and 1 (not include 0 and 1) that represents the proportion of subjects that are assigned with an exposure
rateC	Rate of the exponential distribution to generate censoring times
min.futime	A numeric value that represents minimum follow-up time (in months). The default value is 0, which means no minimum follow-up time is considered. If it has a positive value, this argument will help exclude subjects that only spend a short amount of time in the study
min.postexp.futime	A numeric value that represents minimum post-exposure follow-up time (in months). The default value is 0, which means no minimum post-exposure follow-up time is considered. If it has a positive value, this argument will help exclude subjects that only spend a short amount of time in the study after their exposure

Value

A data.frame object with columns corresponding to

id	Integer that represents a subject<e2><80><99>s identification number
start	For counting process formulation. Represents the start of each time interval
stop	For counting process formulation. Represents the end of each time interval
status	Indicator of event. status = 1 when event occurs and 0 otherwise
x	Indicator of exposure. x = 1 when exposed and 0 otherwise

Author(s)

Danyi Xiong, Teeranan Pokaprakarn, Nusrat Rabbee
Maintainer: Nusrat Rabbee <rabbee@berkeley.edu>

Examples

```
sim_data <- tdSim.method2(500, duration=24,lambda12=1.3,
lambda23=0.04, lambda13=0.03, exp.prop=0.2,rateC=0.05,
min.futime=4, min.postexp.futime=4)
```

Index