Package 'longsurr'

November 9, 2025

Type Package

Title Longitudinal Surrogate Marker Analysis

Version 1.1

Description Assess the proportion of treatment effect explained by a longitudinal surrogate marker as described in Agniel D and Parast L (2021) <doi:10.1111/biom.13310>; and estimate the treatment effect on a longitudinal surrogate marker as described in Wang et al. (2025) <doi:10.1093/biomtc/ujaf104>. A tutorial for this package can be found at <https://www.laylaparast.com/longsurr>.

License GPL

Imports stringr, splines, mgcv, Rsurrogate, dplyr, here, tidyr, fs, KernSmooth, stats, fdapace, grf, lme4, mvnfast, plyr, tibble, magrittr, glue, purrr, readr, refund, fda, fda.usc, survival, MASS

NeedsCompilation no

Author Layla Parast [aut, cre], Denis Agniel [aut], Xuan Wang [aut]

Maintainer Layla Parast <parast@austin.utexas.edu>

Depends R (>= 3.5.0)

Repository CRAN

Date/Publication 2025-11-09 12:40:02 UTC

Contents

data_sjm *Example data for semiparametric joint estimation functions*

Description

Simulated example data for semiparametric joint estimation functions

Usage

```
data("data_sjm")
```

Format

A list with 200 observations on the following:

`delta` numeric vector containing the event indicator for each observation

`obsT` numeric matrix containing the time that the surrogate marker was measured for each observation; number of rows is equal to the number of observations (200) and number of columns is equal to the maximum number of surrogate markers measured (15)

`Y` numeric matrix containing the surrogate marker measurements over time for each observation; same dimension as obsT

`Time` numeric vector containing the observed event or censoring time for each observation

`Treatment` numeric vector containing the treatment indicator for each observation with 1 for treated and 0 for control

Examples

```
data(data_sjm)
names(data_sjm)
```

estimate_surrogate_value *Estimate the surrogate value of a longitudinal marker*

Description

Estimate the surrogate value of a longitudinal marker

Usage

```
estimate_surrogate_value(y_t, y_c, X_t, X_c, method = c("gam", "linear",
  "kernel"), k = 3, var = FALSE, bootstrap_samples = 50, alpha = 0.05)
```

Arguments

y_t	vector of n1 outcome measurements for treatment group
y_c	vector of n0 outcome measurements for control or reference group
X_t	n1 x T matrix of longitudinal surrogate measurements for treatment group, where T is the number of time points
X_c	n0 x T matrix of longitudinal surrogate measurements for control or reference group, where T is the number of time points
method	method for dimension-reduction of longitudinal surrogate, either 'gam', 'linear', or 'kernel'
k	number of eigenfunctions to use in semimetric
var	logical, if TRUE then standard error estimates and confidence intervals are provided
bootstrap_samples	number of bootstrap samples to use for standard error estimation, used if var = TRUE, default is 50
alpha	alpha level, default is 0.05

Value

a tibble containing estimates of the treatment effect (Deltahat), the residual treatment effect (Deltahat_S), and the proportion of treatment effect explained (R); if var = TRUE, then standard errors of Deltahat_S and R are also provided (Deltahat_S_se and R_se), and quantile-based 95% confidence intervals for Deltahat_S and R are provided (Deltahat_S_ci_l [lower], Deltahat_S_ci_h [upper], R_ci_l [lower], R_ci_u [upper])

References

Agniel D and Parast L (2021). Evaluation of Longitudinal Surrogate Markers. Biometrics, 77(2): 477-489.

Examples

```
library(dplyr)
data(full_data)


wide_ds <- full_data %>%
dplyr::select(id, a, tt, x, y) %>%
tidyr::spread(tt, x)

wide_ds_0 <- wide_ds %>% filter(a == 0)
wide_ds_1 <- wide_ds %>% filter(a == 1)
X_t <- wide_ds_1 %>% dplyr::select(`-1`:`1`) %>% as.matrix
y_t <- wide_ds_1 %>% pull(y)
X_c <- wide_ds_0 %>% dplyr::select(`-1`:`1`) %>% as.matrix
y_c <- wide_ds_0 %>% pull(y)

estimate_surrogate_value(y_t = y_t, y_c = y_c, X_t = X_t, X_c = X_c,
```

```
method = 'gam', var = FALSE)
estimate_surrogate_value(y_t = y_t, y_c = y_c, X_t = X_t, X_c = X_c,
method = 'linear', var = TRUE, bootstrap_sample = 50)
```

`full_data` *Example data to illustrate functions*

Description

Simulated nonsmooth data to illustrate functions

Usage

```
data("full_data")
```

Format

A data frame with 10100 observations on the following 5 variables.

`id` a unique person ID

`a` treatment group, 0 or 1

`tt` time

`x` surrogate marker value

`y` primary outcome

`presmooth_data` *Pre-smooth sparse longitudinal data*

Description

Pre-smooth sparse longitudinal data

Usage

```
presmooth_data(obs_data, ...)
```

Arguments

`obs_data`	data.frame or tibble containing the observed data, with columns `id` identifying the individual measured, `tt` identifying the time of the observation, `x` the value of the surrogate at time `tt`, and `a` indicating 1 for treatment arm and 0 for control arm.
`...`	additional arguments passed on to `fpca`

Value

list containing matrices `X_t` and `X_c`, which are the smoothed surrogate values for the treated and control groups, respectively, for use in downstream analyses

Examples

```
library(dplyr)
data(full_data)
obs_ds <- group_by(full_data, id)
obs_data <- sample_n(obs_ds, 5)
obs_data <- ungroup(obs_data)

head(obs_data)
presmooth_X <- presmooth_data(obs_data)
```

---

sjm_linear_estimate *Semiparametric Joint Modeling of the Treatment Effect on a Longitudinal Surrogate with a Linear Model*

Description

Semiparametric joint modeling of the treatment effect on a longitudinal surrogate using both a Cox proportional hazards model and linear model

Usage

```
sjm_linear_estimate(X, Time, Delta, obsT, Y, n.resample=100, var = FALSE)
```

Arguments

`X`	numeric vector containing the treatment indicator for each observation with 1 for treated and 0 for control
`Time`	numeric vector containing the observed event or censoring time for each observation
`Delta`	numeric vector containing the event indicator for each observation
`obsT`	numeric matrix containing the time that the surrogate marker was measured for each observation; number of rows should be equal to the number of observations and number of columns should be equal to the maximum number of surrogate markers measured. If the surrogate marker was not measured, the corresponding entry should be 0 or NA.
`Y`	numeric matrix containing the the surrogate marker measurements over time for each observation; number of rows should be equal to the number of observations and number of columns should be equal to the maximum number of surrogate markers measured. If the surrogate marker was not measured, as determined by the obsT entry, the Y at that time will be ignored.
`n.resample`	number of resampled estimates used for variance estimation; default is 100.
`var`	logical indicating whether the user would like variance estimates and confidence intervals; default is FALSE.

Value

A list of estimates is returned:

est	vector of point estimates where the first entry is the hazard ratio from the Cox model, the second entry is the estimated treatment effect on the surrogate marker at baseline, and the third entry is the estimated treatment on the slope of the surrogate marker i.e., the surrogate marker trajectory
SE	if var is TRUE, a vector of standard error estimates corresponding to the returned point estimates
CI_lower	if var is TRUE, a vector of estimates for the lower bound of the 95% confidence interval for the quantities corresponding to the returned point estimates
CI_upper	if var is TRUE, a vector of estimates for the upper bound of the 95% confidence interval for the quantities corresponding to the returned point estimates

Author(s)

Xuan Wang

References

Wang X, Zhou J, Parast L, Greene T (2025). Semiparametric Joint Modeling to Estimate the Treatment Effect on a Longitudinal Surrogate with Application to Chronic Kidney Disease Trials. Biometrics, 81(3): ujaf104.

Examples

```
data(data_sjm)

sjm_linear_estimate(X=data_sjm$Treatment, Time = data_sjm$Time,
Delta = data_sjm$delta, obsT = data_sjm$obsT, Y = data_sjm$Y)

sjm_linear_estimate(X=data_sjm$Treatment, Time =
data_sjm$Time, Delta = data_sjm$delta, obsT = data_sjm$obsT,
Y = data_sjm$Y, n.resample=5, var=TRUE)
```

---

sjm_nl_estimate *Semiparametric Joint Modeling of the Treatment Effect on a Longitudinal Surrogate with a Nonlinear Model*

Description

Semiparametric joint modeling of the treatment effect on a longitudinal surrogate using both a Cox proportional hazards model and a splines-based model

Usage

```
sjm_nl_estimate(X, Time, Delta, obsT, Y, gap_time = 0.1, n.resample = 100,
var = FALSE)
```

Arguments

X	numeric vector containing the treatment indicator for each observation with 1 for treated and 0 for control
Time	numeric vector containing the observed event or censoring time for each observation
Delta	numeric vector containing the event indicator for each observation
obsT	numeric matrix containing the time that the surrogate marker was measured for each observation; number of rows should be equal to the number of observations and number of columns should be equal to the maximum number of surrogate markers measured. If the surrogate marker was not measured, the corresponding entry should be 0 or NA.
Y	numeric matrix containing the the surrogate marker measurements over time for each observation; number of rows should be equal to the number of observations and number of columns should be equal to the maximum number of surrogate markers measured. If the surrogate marker was not measured, as determined by the obsT entry, the Y at that time will be ignored.
gap_time	number indicating gap time for slope estimation; default is 0.1.
n.resample	number of resampled estimates used for variance estimation; default is 100.
var	logical indicating whether the user would like variance estimates and confidence intervals; default is FALSE.

Value

A list of estimates is returned:

est	estimated hazard ratio from the Cox model
est_t	vector of estimated treatment effect on the slope of the surrogate marker i.e., the surrogate marker trajectory, on a grid constructed from the given gap time
t_grid	vector of grid times corresponding to the returned estimates
SE_est	if var is TRUE, standard error estimate of the hazard ratio
SE_est_t	if var is TRUE, standard error estimate of the estimated treatment effect on the slope of the surrogate marker
CI_lower_est	if var is TRUE, lower bound of the 95% confidence interval for the hazard ratio
CI_upper_est	if var is TRUE, upper bound of the 95% confidence interval for the hazard ratio
CI_lower_est_t	if var is TRUE, lower bound of the 95% confidence interval for the treatment effect on the slope of the surrogate marker
CI_upper_est_t	if var is TRUE, upper bound of the 95% confidence interval for the treatment effect on the slope of the surrogate marker

Author(s)

Xuan Wang

References

Wang X, Zhou J, Parast L, Greene T (2025). Semiparametric Joint Modeling to Estimate the Treatment Effect on a Longitudinal Surrogate with Application to Chronic Kidney Disease Trials. Biometrics, 81(3): ujaf104.

Examples

```
data(data_sjm)


sjm_nl_estimate(X=data_sjm$Treatment, Time = data_sjm$Time,
Delta = data_sjm$delta, obsT = data_sjm$obsT, Y = data_sjm$Y, gap_time=0.2)

sjm_nl_estimate(X=data_sjm$Treatment, Time =
data_sjm$Time, Delta = data_sjm$delta, obsT = data_sjm$obsT,
Y = data_sjm$Y, gap_time = 0.2, n.resample=5, var=TRUE)
```

Index