Package 'survivalsurrogate'

May 2, 2025

Type Package

Title Evaluate a Longitudinal Surrogate with a Censored Outcome

Version 1.0

Date 2025-05-02

Description Provides influence function-based methods to evaluate a longitudinal surrogate marker in a censored time-to-event outcome setting, with plug-in and targeted minimum loss-based estimation options. More details will be available in the future in: Agniel D and Parast L (2025+). ``Robust Evaluation of Longitudinal Surrogate Markers with Censored Data." Journal of the Royal Statistical Society: Series B, In press.

License GPL

Imports stats, dplyr, magrittr, glue, mlr3, purrr, SparseM, rBeta2009, data.table

NeedsCompilation no

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

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

Contents

`estimate_R` *Estimates the proportion of the treatment effect explained*

Description

Estimates the proportion of the treatment effect on the censored primary outcome that is explained by the longitudinal surrogate marker

Usage

```
estimate_R(delta_if, delta_s_if, delta = NULL, delta_s = NULL, se_type = "asymptotic", n_boot = NULL
```

Arguments

delta_if	delta influence function
delta_s_if	delta.s influence function
delta	optional; defaults to the mean of the delta influence function
delta_s	optional; defaults to the mean of the delta.s influence function
se_type	choices are "asymptotic" or "bootstrap"
n_boot	number of boostrap samples, required if se_type = "bootstrap"
alpha	alpha level used for confidence interval, defaults to 95% confidence interval

Value

returns a tibble with the estimate, standard error (se), and confidence interval limits for delta, delta.s, and R (the proportion of treatment effect explained)

References

Agniel D and Parast L (2025+). "Robust Evaluation of Longitudinal Surrogate Markers with Censored Data." Journal of the Royal Statistical Society: Series B, In press.

---

plugin_delta *Plug-in Estimator of Causal Effect Based on EIF*

Description

Estimates a causal parameter using a plug-in estimator that combines machine learning-based nuisance function estimation and efficient influence function (EIF) correction.

Usage

```
plugin_delta(data, folds, id, x, g, a = NULL, y, s, binary_lrnr = NULL,
             cont_lrnr = NULL, e = NULL, gamma1 = NULL, gamma0 = NULL,
             mu1 = NULL, mu0 = NULL, Q1 = NULL, Q0 = NULL,
             Qstar1 = NULL, Qstar0 = NULL, truncate_e = 1e-12,
             verbose = FALSE)
```

Arguments

data	A data.table or data.frame containing the observed data.
folds	A vector or list indicating cross-validation folds, used for sample splitting.
id	The name of the ID variable (as a string).
x	Vector of covariate column names used in nuisance estimation.
g	The name of the treatment indicator variable (as a string).
a	(Optional) Names of indicators for availability or adherence, if relevant.
y	Names of the outcome variables (can be longitudinal, e.g., Y_1, Y_2, ...).

s	Names of survival indicator variables, aligned with y.
binary_lrnr	Learner object used for binary outcome nuisance functions (e.g., treatment, censoring, hazard).
cont_lrnr	Learner object used for continuous outcome nuisance functions.
e	(Optional) Name of estimated propensity score variable. If NULL, it is estimated from the data.
gamma1	(Optional) Names of estimated censoring probabilities under treatment. If NULL, they are estimated.
gamma0	(Optional) Names of estimated censoring probabilities under control. If NULL, they are estimated.
mu1	(Optional) Names of estimated hazards under treatment. If NULL, they are estimated.
mu0	(Optional) Names of estimated hazards under control. If NULL, they are estimated.
Q1	(Optional) Names of estimated conditional means under treatment. If NULL, they are estimated.
Q0	(Optional) Names of estimated conditional means under control. If NULL, they are estimated.
Qstar1	(Optional) Placeholder for TMLE-style fluctuation targets (currently unused).
Qstar0	(Optional) Placeholder for TMLE-style fluctuation targets (currently unused).
truncate_e	Threshold for truncating small propensity scores to stabilize estimation.
verbose	Logical; if TRUE, prints messages during estimation.

Details

This function estimates a target causal effect using a plug-in approach based on nuisance functions: hazard, censoring, and outcome models. It also constructs the efficient influence function (EIF) and returns an estimate, its standard error, and the full EIF sample.

Value

A data.table with the following components:

plugin_est	Estimated causal effect.
plugin_se	Estimated standard error based on the EIF.
if_data	A nested data.table containing the EIF values for each observation.

References

Agniel D and Parast L (2025+). "Robust Evaluation of Longitudinal Surrogate Markers with Censored Data." Journal of the Royal Statistical Society: Series B, In press.

plugin_delta_s *Plugin Estimator of the Surrogate Causal Effect at a Fixed Time Point*

Description

Computes a plugin estimator of the average causal effect of treatment on a primary outcome, conditionally mediated through a surrogate marker, at a fixed time point. This estimator uses flexible machine learning models for nuisance parameters and supports both asymptotic and bootstrap standard errors.

Usage

```
plugin_delta_s(data, folds, id, x, g, a = NULL, y, s, binary_lrnr = NULL, cont_lrnr = NULL, t0 = lengt
```

Arguments

data	A data.table or data.frame with one row per observation.
folds	A vector of fold assignments used for cross-fitting.
id	The name of the column identifying individual observations.
x	A character vector of baseline covariate column names.
g	The name of the treatment assignment column (0 or 1).
a	Optional; vector of censoring indicator column names at each time point. If NULL, assumes no censoring.
y	Vector of primary outcome indicator column names at each time point.
s	Vector of surrogate marker indicator column names at each time point.
binary_lrnr	A binary outcome learner (e.g., a classification model) used to estimate binary nuisance parameters such as hazards and censoring probabilities.
cont_lrnr	A continuous outcome learner (e.g., regression model) used to estimate conditional expectations.
t0	The time index at which the causal effect is evaluated. Defaults to the last time point.
e	Optional; the name of the column containing estimated treatment propensity scores. If NULL, they are estimated.
gamma1	Optional; vector of column names for estimated censoring probabilities under treatment. If NULL, they are estimated or assumed to be 1.
gamma0	Optional; vector of column names for estimated censoring probabilities under control. If NULL, they are estimated or assumed to be 1.
mu1	Optional; vector of column names for hazards under treatment. If NULL, they are estimated.
mu0	Optional; vector of column names for hazards under control. If NULL, they are estimated.
pi	Optional; vector of column names for estimated probabilities of surrogate marker under control. If NULL, they are estimated.
pistar	Optional; vector of column names for estimated probabilities of surrogate marker under treatment. If NULL, they are estimated.

Q1	Optional; vector of column names for conditional expectations of outcome under treatment. If NULL, they are estimated.
Q0	Optional; vector of column names for conditional expectations of outcome under control. If NULL, they are estimated.
Qstar1	Optional; vector of column names for pseudo-outcome expectations under treatment. If NULL, they are estimated.
Qstar0	Optional; vector of column names for pseudo-outcome expectations under control. If NULL, they are estimated.
truncate_e	Truncation threshold for propensity scores to avoid division by zero. Default is 1e-12.
truncate_pi	Truncation threshold for surrogate probabilities to avoid instability. Default is 1e-12.
se_type	Method to estimate standard errors: either "asymptotic" or "bootstrap".
n_boot	Number of bootstrap replicates if se_type = "bootstrap". Required if using bootstrap.
alpha	Significance level for confidence intervals. Default is 0.05.
verbose	Logical; if TRUE, prints progress messages.
retain_data	Logical; if TRUE, retains full dataset with estimated components in the output.

Details

This function computes a plug-in estimator for the surrogate causal effect, i.e., the average causal effect of treatment on the primary outcome that is mediated by a surrogate marker, at a specific time point. The function optionally estimates nuisance parameters (e.g., hazards, expectations, surrogate probabilities) using cross-fitting and machine learning, and returns estimates with standard errors and confidence intervals.

Value

A data.frame with the following components:

plugin_est	The plug-in point estimate of the surrogate causal effect at time t0.
plugin_se	Standard error of the estimate, based on influence function or bootstrap.
ci_l	Lower bound of the (1 - alpha) confidence interval.
ci_h	Upper bound of the (1 - alpha) confidence interval.
if_data	A list containing the dataset with the efficient influence function if retain_data = TRUE.

References

Agniel D and Parast L (2025+). "Robust Evaluation of Longitudinal Surrogate Markers with Censored Data." Journal of the Royal Statistical Society: Series B, In press.

`tmle_delta` *Targeted Maximum Likelihood Estimation with Censoring Adjustment*

Description

Estimates a causal effect using Targeted Maximum Likelihood Estimation (TMLE) while accounting for time-dependent outcomes and potential censoring under treatment and control. This implementation supports flexible machine learning-based nuisance parameter estimation.

Usage

```
tmle_delta(data, folds, id, x, g, a = NULL, y, s, binary_lrnr = NULL, cont_lrnr = NULL, e = NULL, gamm
```

Arguments

`data`	A `data.frame` containing all relevant variables for estimation, including covariates, treatment, censoring, and outcome variables.
`folds`	An object specifying the cross-validation folds to use for nuisance parameter estimation. Typically a list of fold assignments.
`id`	A character string giving the name of the column identifying individual units in the data.
`x`	A character vector specifying the baseline covariates to adjust for.
`g`	A character string giving the name of the treatment variable (binary indicator for treatment vs. control).
`a`	(Optional) A character vector of names for the time-varying treatment indicators. If `NULL`, assumes no censoring.
`y`	A character vector of names for the outcome variables at each time point.
`s`	A character vector of names for the indicators of being at risk (uncensored) at each time point.
`binary_lrnr`	A binary learner (e.g., SuperLearner object or algorithm name) used to estimate the propensity score and censoring mechanisms.
`cont_lrnr`	A continuous outcome learner used to estimate the conditional expectations of the outcomes.
`e`	(Optional) A character string giving the name of the column with known or externally estimated treatment propensity scores. If `NULL`, these are estimated from the data.
`gamma1`	(Optional) A character vector of names of columns giving the censoring probabilities under treatment. If `NULL`, these are estimated.
`gamma0`	(Optional) A character vector of names of columns giving the censoring probabilities under control. If `NULL`, these are estimated.
`truncate_e`	A numeric value indicating the truncation level for the propensity score, used to avoid division by near-zero values.
`verbose`	Logical; if `TRUE`, prints progress and estimation details.

Details

This function estimates the average treatment effect on a time-indexed outcome using TMLE, incorporating estimates of the treatment assignment mechanism, censoring processes, and outcome regressions. It supports estimation under censoring by estimating or supplying the censoring probabilities under both treatment and control. The user can optionally supply known values for the propensity score and censoring probabilities.

Value

A data frame with the following components:

`tmle_est`	The estimated causal effect (average difference in potential outcomes).
`tmle_se`	The estimated standard error of the TMLE estimator, based on the influence function.
`if_data`	A nested data frame containing the influence function contributions for each observation.

References

Agniel D and Parast L (2025+). "Robust Evaluation of Longitudinal Surrogate Markers with Censored Data." Journal of the Royal Statistical Society: Series B, In press.

`tmle_delta_s` *TMLE for Longitudinal Surrogate Evaluation with Delta Statistic*

Description

Estimates a robust treatment effect using targeted minimum loss-based estimation (TMLE) for evaluating longitudinal surrogate markers under potential censoring. The method computes a delta-type estimand based on the difference between counterfactual outcomes, using influence-function-based standard errors.

Usage

```
tmle_delta_s(data, folds, id, x, g, a = NULL, y, s, binary_lrnr = NULL, cont_lrnr = NULL, t0 = length(
```

Arguments

`data`	A `data.frame` containing all necessary variables for estimation.
`folds`	A vector or list defining cross-validation folds for nuisance estimation.
`id`	A string giving the name of the column with unique unit IDs (e.g., subject ID).
`x`	A character vector of covariate names to adjust for in nuisance estimation.
`g`	A string indicating the column name of the treatment indicator (typically 0/1).
`a`	(Optional) A character vector of time-varying treatment columns. If `NULL`, assumes static treatment.
`y`	A character vector of outcome variable names (possibly longitudinal).
`s`	A character vector of surrogate marker variable names (possibly longitudinal).
`binary_lrnr`	Learner object or specification used for estimating binary nuisance components (e.g., propensity scores, censoring).

cont_lrnr	Learner object used for estimating continuous-valued outcome regressions.
t0	Time index (default is the final time point) at which the delta estimand is evaluated.
e	(Optional) Column name or vector of propensity score estimates. If NULL, these will be estimated.
gamma1	(Optional) Vector of column names for censoring probabilities under treatment. Estimated if NULL.
gamma0	(Optional) Vector of column names for censoring probabilities under control. Estimated if NULL.
pi	(Optional) Vector of column names giving estimated probabilities of observed surrogate values under observed treatment. Estimated if NULL.
pistar	(Optional) Vector of column names giving estimated probabilities under the reference distribution. Estimated if NULL.
truncate_e	Numeric truncation level for propensity scores to avoid division by near-zero values. Default is 1e-12.
verbose	Logical; if TRUE, print progress messages.
retain_data	Logical; if TRUE, return the full data and influence function values.

Value

A data.frame with the following components:

tmle_est	TMLE estimate of the treatment effect based on surrogate delta statistic.
tmle_se	Estimated standard error using the empirical standard deviation of the influence function.
if_data	(Optional) Full dataset joined with estimated influence function contributions.

References

Agniel D and Parast L (2025+). "Robust Evaluation of Longitudinal Surrogate Markers with Censored Data." *Journal of the Royal Statistical Society: Series B*, In press.

Index