

Package ‘DRsurvCRT’

December 30, 2025

Type Package

Title Doubly-Robust Estimation for Survival Outcomes in Cluster-Randomized Trials

Version 0.0.1

Description Cluster-randomized trials (CRTs) assign treatment to groups rather than individuals, so valid analyses must distinguish cluster-level and individual-level effects and define estimands within a potential-outcomes framework. This package supports right-censored survival outcomes for both single-state (binary) and multi-state settings. For single-state outcomes, it provides estimands based on stage-specific survival contrasts (SPCE) and restricted mean survival time (RMST). For multi-state outcomes, it provides SPCE as well as a generalized win-based restricted mean time-in-favor estimand (RMT-IF). The package implements doubly robust estimators that accommodate covariate-dependent censoring and remain consistent if either the outcome model or the censoring model is correctly specified. Users can choose marginal Cox or gamma-frailty Cox working models for nuisance estimation, and inference is supported via leave-one-cluster-out jackknife variance and confidence interval estimation. Methods are described in Fang et al. (2025) ``Estimands and doubly robust estimation for cluster-randomized trials with survival outcomes'' <[doi:10.48550/arXiv.2510.08438](https://doi.org/10.48550/arXiv.2510.08438)>.

License MIT + file LICENSE

Encoding UTF-8

LazyData true

Imports Rcpp, frailtyEM, survival, ggplot2, pracma, abind

LinkingTo Rcpp, RcppArmadillo

RoxgenNote 7.3.3

Depends R (>= 3.5)

NeedsCompilation yes

Author Xi Fang [aut, cre],
Fan Li [aut]

Maintainer Xi Fang <x.fang@yale.edu>

Repository CRAN

Date/Publication 2025-12-30 19:00:06 UTC

Contents

| | |
|-----------------------------|---|
| datm | 2 |
| dats | 3 |
| DRsurvfit | 5 |
| plot.DRsurvfit | 8 |
| summary.DRsurvfit | 9 |

Index

11

| | |
|------|--|
| datm | <i>Simulated multi-state CRT survival data (long format)</i> |
|------|--|

Description

A simulated cluster-randomized trial (CRT) dataset with a *multi-state* prioritized endpoint and right censoring, stored in long format. Each individual can experience up to several ordered event types, and contributes multiple rows with increasing time. The data are designed to illustrate the multi-state SPCE and RMT-IF estimators in **DRsurvCRT**.

Usage

datm

Format

A data frame with N rows and 9 variables:

id Individual identifier. Each subject may appear on multiple rows.

time Observed time of the record (e.g., transition time or censoring time). Times are non-decreasing within an id.

event Multi-state event indicator. The coding is:

- 0 = still at-risk / no new transition (often the last row for a censored subject);
- 1, 2, 3 = entry into progressively more severe event states (e.g., non-fatal event, progression, death), in increasing clinical priority.

Not every subject experiences all states; some may be censored earlier.

cluster Integer cluster identifier.

trt Cluster-level treatment indicator; 0 = control, 1 = intervention. Constant within cluster.

W1 Cluster-level baseline binary covariate.

W2 Cluster-level baseline continuous covariate.

Z1 Individual-level continuous baseline covariate.

Z2 Individual-level binary baseline covariate.

Details

For each subject id, rows are ordered by time. If an individual transitions to a new state (e.g., from state 0 to 1, or from 1 to 2, etc.), the corresponding transition time is recorded with event > 0 . Later rows for the same subject may reflect further transitions or censoring.

This structure is suitable for constructing state-specific transformed survival functions $S_{a,s}(t)$ (for treatment arm a and state s) and for computing stage-wise RMT-IF contributions for prioritized composite endpoints. It is the canonical example dataset used by the multi-state core and jackknife routines in **DRsurvCRT**.

Examples

```
data(datm)

## quick look
head(datm)
table(datm$cluster, datm$trt)
table(datm$event)

## Example: fit multi-state SPCE / RMT-IF (schematic)

fit_ms <- DRsurvfit(
  data      = datm,
  formula   = survival::Surv(time, event) ~ W1 + W2 + Z1 + Z2 + cluster(cluster),
  intv     = "trt",
  method    = "frailty",
  estimand  = "SPCE",           # or "RMTIF"
  variance  = "jackknife"
)

summary(fit_ms, level = "cluster")
plot(fit_ms, level = "cluster")
```

data

Simulated single-state CRT survival data

Description

A simulated cluster-randomized trial (CRT) dataset with a *single* terminal event type and right censoring. Each row corresponds to one individual in a cluster. The data are intended for illustrating the doubly-robust survival estimators in **DRsurvCRT** for the simple two-state setting (alive vs. terminal event).

Usage

data

Format

A data frame with n rows and 8 variables:

cluster Integer cluster identifier.

id Integer individual identifier within cluster.

trt Cluster-level treatment indicator; 0 = control, 1 = intervention. Constant within cluster.

W1 Cluster-level baseline binary covariate (e.g., center-level characteristic).

W2 Cluster-level baseline continuous covariate.

Z1 Individual-level continuous baseline covariate.

Z2 Individual-level binary baseline covariate.

time Observed follow-up time (event or censoring time).

event Event indicator; 1 = terminal event, 0 = right censored.

Details

The dataset was generated from a cluster-randomized design with covariate-dependent hazards and administrative censoring. It is primarily used to demonstrate calls such as `DRsurvfit(..., estimand = "SPCE")` and the corresponding variance and plotting methods in the single-state setting.

Examples

```
data(datas)

## quick look
head(datas)
table(datas$cluster, datas$trt)

## Example use with DRsurvfit (marginal Cox working model)

fit_spce <- DRsurvfit(
  data    = datas,
  formula = survival::Surv(time, event) ~ W1 + W2 + Z1 + Z2 + cluster(cluster),
  intv   = "trt",
  method  = "marginal",
  estimand = "SPCE",
  variance = "jackknife"
)

summary(fit_spce)
plot(fit_spce, level = "cluster")
```

Description

Fits doubly-robust estimators for cluster-randomized trials with right-censored survival outcomes, including single-state and multi-state outcomes

The outcome is specified as `Surv(time, status)`, where `status` in $\{0, 1, 2, \dots, Q\}$ and `status = 0` denotes censoring. Values $1, 2, \dots, Q$ are ordered states, with the largest state typically representing an absorbing state (e.g., death).

The function supports two estimands:

- **SPCE**: stage-specific survival probabilities $S_s(t)$ for each state $s = 1, \dots, S_{\max}$ at all event times.
- **RMTIF**: a generalized win-based restricted mean time in favor estimand constructed from the multi-state survival outcome. When `status` is binary (0/1), this reduces to an RMST estimand (evaluated on the full observed-time grid).
- **RMST**: a special case of **RMTIF** when `status` in $\{0, 1\}$ (one nonzero state). In this case the generalized RMT-IF reduces to a regular RMST contrast.

Jackknife variance is computed via leave-one-cluster-out re-fitting method

- For `estimand = "SPCE"`: variances of $S_1(t)$, $S_0(t)$, and $S_1(t) - S_0(t)$ at each time and state.
- For `estimand = "RMTIF"`: variances and covariance of $R_1(\tau)$, $R_0(\tau)$, and $R_1(\tau) - R_0(\tau)$ at each event time τ .

The returned object includes metadata needed for summaries and plotting: final fitted outcome/censoring formulas, the cluster id column, number of clusters, degrees of freedom for jackknife t-intervals ($= M - 1$), sample sizes, and the cluster-level and individual-level estimators.

Usage

```
DRsurvfit(
  data,
  formula,
  cens_formula = NULL,
  intv,
  id_var = NULL,
  method = c("marginal", "frailty"),
  estimand = c("SPCE", "RMTIF", "RMST"),
  trt_prob = NULL,
  variance = c("none", "jackknife"),
  fit_controls = NULL,
  verbose = FALSE
)
```

Arguments

| | |
|---------------------------|---|
| <code>data</code> | A <code>data.frame</code> . |
| <code>formula</code> | Outcome model: e.g., <code>Surv(time, status) ~ W1 + W2 + Z1 + Z2 + cluster(M)</code> . The left-hand side must be <code>Surv(time, status)</code> with <code>status</code> in $\{0, 1, 2, \dots\}$ and 0 indicating censoring. The right-hand side <i>must</i> include a <code>cluster(<id>)</code> term specifying the cluster id for CRTs. All other covariates may be individual- or cluster-level. |
| <code>cens_formula</code> | Optional censoring model. If <code>NULL</code> , the censoring model is built automatically from the outcome formula by: <ul style="list-style-type: none"> • reusing the RHS (excluding <code>cluster()</code>); • using LHS <code>Surv(time, event == 0)</code>; If supplied, <code>cens_formula</code> is used as-is for all stage-specific fits, but the DR estimating equations still use the stage-specific event indicator as described above. |
| <code>intv</code> | Character: name of the cluster-level treatment column (0/1), constant within cluster. |
| <code>id_var</code> | Character: name of the individual id column. If <code>NULL</code> , considered as single state. |
| <code>method</code> | "marginal" or "frailty". <ul style="list-style-type: none"> • "marginal": fits <code>survival::coxph</code> models with <code>cluster(<id>)</code> robust variance. • "frailty": fits <code>frailtyEM::emfrail</code> gamma-frailty models for outcome and censoring. |
| <code>estimand</code> | "SPCE", "RMTIF", or "RMST". <ul style="list-style-type: none"> • "SPCE": returns stage-specific survival arrays <code>S_stage_cluster</code> and <code>S_stage_ind</code> with dimensions $[time \times 2 \times Q]$. • "RMTIF": returns the generalized win-based restricted mean time in favor estimand at each event time, along with stage-wise contributions. For a binary status, this reduces to an RMST estimands. • "RMST": restricted mean survival time difference for the binary case <code>status</code> in $\{0, 1\}$. This is a convenience alias for the "RMTIF" calculations when there is exactly one nonzero event state. |
| <code>trt_prob</code> | Optional length-2 numeric vector (<code>p0, p1</code>) giving the cluster-level treatment probabilities for arms 0 and 1. If <code>NULL</code> , they are computed as the empirical proportion of treatment assignments per cluster. |
| <code>variance</code> | "none" or "jackknife" for variance estimation. |
| <code>fit_controls</code> | Optional <code>frailtyEM::emfrail_control()</code> list, used only when <code>method = "frailty"</code> . If <code>NULL</code> , default fast-fitting controls are used (no standard errors from the <code>frailtyEM</code> fits are required here). |
| <code>verbose</code> | Logical; currently unused but kept for future verbosity options. |

Value

An object of class "DRsurvfit" with fields depending on estimand:

- Common:**
- method: fitted method ("marginal" or "frailty").
 - estimand: requested estimand ("SPCE" or "RMTIF").
 - trt_prob: numeric vector $c(p_0, p_1)$.
 - event_time: time grid:
 - SPCE: all event times including 0.
 - RMTIF: positive event times τ at which the RMT-IF is evaluated.
 - max_state: maximum observed non-zero status.
 - cluster_col: name of the cluster id column.
 - n_clusters: number of clusters (M).
 - df_jackknife: jackknife degrees of freedom ($M - 1$).
 - n_obs: total number of observations.
 - n_events: total number of non-censoring observations (status != 0).
 - cluster_trt_counts: counts of treated and control clusters $c(n_{trt0}, n_{trt1})$ based on first row per cluster.
 - formula_outcome: fully reconstructed outcome formula.
 - cens_formula: final censoring formula used.
 - call: the matched call.
 - jackknife: logical indicating whether jackknife variance was computed.
- If estimand = "SPCE":**
- S_stage_cluster: 3D array [$time \times 2 \times Q$] with stage-specific cluster-level survival: $S_{stage_cluster[, 1, s]} = S_1^{(s)}(t)$, $S_{stage_cluster[, 2, s]} = S_0^{(s)}(t)$.
 - S_stage_ind: analogous individual-level survival array.
 - var_stage_cluster: jackknife variances for $S_1^{(s)}(t)$, $S_0^{(s)}(t)$, and $S_1^{(s)}(t) - S_0^{(s)}(t)$ as a 3D array [$time \times 3 \times Q$] with dimension names $comp = c("Var(S1)", "Var(S0)", "Var(S1-S0)")$, when variance = "jackknife"; otherwise NULL.
 - var_stage_ind: analogous individual-level variance array.
- If estimand = "RMTIF":**
- RMTIF_cluster: matrix [$time \times 3$] with columns $c("R1", "R0", "R1-R0")$ giving the cluster-level RMT-IF curves at each event time τ .
 - RMTIF_ind: analogous individual-level RMT-IF matrix.
 - stagewise_cluster: list of length $length(event_time)$; each element is a $3 \times (Q)$ matrix of stage-wise contributions with rows $c("s1qs0qp1", "s0qs1qp1", "diff")$ and columns $c("stage_1", ..., "stage_Q", "sum")$.
 - stagewise_ind: analogous individual-level list.
 - var_rmtif_cluster: jackknife variance/covariance matrix [$time \times 4$] with columns $c("Var(R1)", "Var(R0)", "Var(R1-R0)", "Cov(R1, R0)")$, when variance = "jackknife"; otherwise NULL.
 - var_rmtif_ind: analogous individual-level matrix.
 - S_stage_cluster, S_stage_ind: the underlying stage-specific survival arrays are also returned for convenience.

Examples

```
data(datm)

## Multi-state RMT-IF (binary reduces to RMST-type)
fit_rmtif <- DRsurvfit(
  datm,
  Surv(time, event) ~ W1 + W2 + Z1 + Z2 + cluster(cluster),
  intv   = "trt",
  method  = "marginal",
  estimand = "RMTIF",
  variance = "none"
)
```

plot.DRsurvfit

Plot method for DRsurvfit objects (SPCE / RMT-IF)

Description

Produces plots for multi-state doubly-robust estimators:

- For `estimand = "SPCE"`: for each state s , plots the difference curve $S_{1,s}(t) - S_{0,s}(t)$ with jackknife t-based confidence bands over time.
- For `estimand = "RMTIF"`: plots the overall RMT-IF difference curve $R_1(t) - R_0(t)$ (sum of stage-wise contributions) with jackknife t-based confidence bands over time.

The argument `tau` is a truncation time: if supplied, the plot is restricted to `event_time <= tau`. If `tau` is `NULL`, the full event-time grid is used.

Usage

```
## S3 method for class 'DRsurvfit'
plot(
  x,
  level = c("cluster", "individual"),
  states = NULL,
  tau = NULL,
  alpha = 0.05,
  ...
)
```

Arguments

| | |
|---------------------|---|
| <code>x</code> | A <code>DRsurvfit</code> object. |
| <code>level</code> | Character: <code>"cluster"</code> or <code>"individual"</code> . |
| <code>states</code> | Optional integer vector of states to plot when <code>estimand = "SPCE"</code> . Defaults to all states <code>1:object\$max_state</code> . |

| | |
|-------|--|
| tau | Optional numeric truncation time. If non-NULL, only event times $\leq \max(\text{tau})$ are plotted. If NULL, all event times are plotted. |
| alpha | Nominal type I error for the intervals; coverage is $1 - \alpha$. Default is 0.05. |
| ... | Unused; included for S3 consistency. |

Value

The input object `x`, invisibly.

summary.DRsurvfit

Summary method for DRsurvfit objects (multi-state SPCE / RMTIF/RMST)

Description

Produces tabular summaries for multi-state doubly-robust estimators:

- For `estimand = "SPCE"`: stage-specific survival probabilities $S_s(t)$ at selected times τ , with optional jackknife t-based confidence intervals.
- For `estimand = "RMTIF"` or `"RMST"`: curves $R_1(\tau)$, $R_0(\tau)$, and $R_1(\tau) - R_0(\tau)$ at the same set of τ , again with optional jackknife t-based intervals.

The same argument `tau` is used for both estimands. If `tau` is NULL, the function uses the 25%, 50%, and 75% quantiles of the event-time grid (excluding time 0 if present).

Usage

```
## S3 method for class 'DRsurvfit'
summary(
  object,
  level = c("cluster", "individual"),
  tau = NULL,
  states = NULL,
  digits = 4,
  alpha = 0.05,
  ...
)
```

Arguments

| | |
|--------|--|
| object | A <code>DRsurvfit</code> object. |
| level | Character: "cluster" or "individual" level summary. |
| tau | Optional numeric vector of times at which to summarize both SPCE and RMTIF/RMST. If NULL, the 25%, 50%, and 75% quantiles of the event-time grid are used. |
| states | Optional integer vector of states to summarize for <code>estimand = "SPCE"</code> . Defaults to all states 1: <code>object\$max_state</code> . |

| | |
|---------------------|--|
| <code>digits</code> | Number of digits to print for estimates and confidence limits. |
| <code>alpha</code> | Nominal type I error for the intervals; coverage is $1 - \text{alpha}$. Default is <code>0.05</code> , giving 95% confidence intervals. |
| <code>...</code> | Additional arguments passed to or from methods. Currently unused. |

Value

The input object `object`, invisibly.

Index

* **datasets**

datm, [2](#)

dats, [3](#)

datm, [2](#)

dats, [3](#)

DRsurvfit, [5](#)

plot.DRsurvfit, [8](#)

summary.DRsurvfit, [9](#)