

Package: drsurv (via r-universe)

September 16, 2024

Type Package

Title Doubly Robust Estimation of Survival Differences with Censored Data

Version 0.1.0

URL <https://github.com/sachsmc/drsurv>

BugReports <https://github.com/sachsmc/drsurv/issues>

Date 2024-01-28

Maintainer Michael C Sachs <sachsmc@gmail.com>

Description An implementation of several doubly robust estimators for the survival difference at a given time point and one more complex doubly robust estimator for the survival curve process. The estimators are doubly robust in the sense that they are consistent if the censoring model is correctly specified for censoring and either the outcome model is correctly specified for confounding or the exposure model is correctly specified for confounding. See <[doi:10.48550/arXiv.2310.16207](https://doi.org/10.48550/arXiv.2310.16207)> for more details and examples.

License MIT + file LICENSE

Encoding UTF-8

LazyData true

Imports numDeriv, boot, mets, timereg, eventglm, stats

RoxygenNote 7.2.3

Depends R (>= 2.10), survival

Suggests knitr, rmarkdown

VignetteBuilder knitr

Repository <https://sachsmc.r-universe.dev>

RemoteUrl <https://github.com/sachsmc/drsurv>

RemoteRef HEAD

RemoteSha 9496887f5bb266318eb60de8a1d0ba360f64c5c7

Contents

blanche_survdiff	2
pseudo_survdiff	4
rotterdam	6
sjovan_survdiff	7

Index	10
--------------	-----------

blanche_survdiff	<i>Estimate the survival probabilities in two treatment groups at several times</i>
------------------	---

Description

Doubly robust estimation of the survival probabilities in two exposure groups and their difference using the method of Blanche et al 2023. By default it is assumed that censoring is completely independent, but covariate-dependence can be modeled by either specifying the right side of `cformula` which will be passed to a Cox model for censoring, or estimating censoring weights externally and passing them to the `cens.weights` argument.

Usage

```
blanche_survdiff(
  oformula,
  cformula = ~1,
  cens.weights = NULL,
  eformula,
  times,
  data,
  weights = NULL,
  subset = NULL
)
```

Arguments

<code>oformula</code>	The outcome formula, the left side should be a call to Event
<code>cformula</code>	The censoring model formula, a one sided formula with no outcome
<code>cens.weights</code>	A vector of inverse probability of censoring weights. If not NULL, then <code>cformula</code> will be ignored.
<code>eformula</code>	The exposure model formula, which will be fit by logistic regression in a call to glm
<code>times</code>	Vector of times at which to estimate the survival probabilities
<code>data</code>	Data frame in which to find the variables in the model formulas
<code>weights</code>	Vector of case weights
<code>subset</code>	Logical vector

Details

As presented in Blanche et al. 2023, we model the causal survival probabilities $p\{T(x) > t\}$. We assume there is a way to correctly specify for confounding a model for $p\{T(x) \leq t|Z\}$ as in a logistic regression model

$$p\{T \leq t|X, \mathbf{Z}\} = Q_t(X, \mathbf{Z}; \beta^\dagger, \gamma^\dagger) = \frac{\exp(\beta^\dagger X + h(X, \mathbf{Z}; \gamma^\dagger))}{1 + \exp(\beta^\dagger X + h(X, \mathbf{Z}; \gamma^\dagger))}.$$

The working propensity model for exposure is a logistic regression model. In addition to these models, we model the censoring distribution as $p\{C > u|X, \mathbf{Z}\} = G_c(u, X, \mathbf{Z})$. If `cens.weights` is not `NULL`, then we estimate $\hat{G}_c(u, X, \mathbf{Z})$ by a Cox model. This can be done externally by Aalen's additive hazard model, nonparametric estimators stratified on covariates, or if censoring is assumed independent of covariates, G_c may be estimated with the Kaplan-Meier estimator or any other nonparametric estimator and passed in the `cens.weights` argument. Let the estimated inverse probability of censoring weights be

$$\hat{U}_i(t) = \frac{I_{\tilde{T}_i \leq t} \Delta_i}{\hat{G}_c(\tilde{T}_i, X_i, \mathbf{Z}_i)} + \frac{I_{\tilde{T}_i > t}}{\hat{G}_c(t, X_i, \mathbf{Z}_i)}.$$

Then let $\hat{\beta}^\dagger, \hat{\gamma}^\dagger$ denote the parameter estimates resulting from the inverse probability of censoring weighted, but propensity score unweighted, score equations for the logistic model estimated via maximum likelihood among the subset of individuals who either had the event before time t or remained under observation until at least time t . Then the doubly robust estimator of the causal event probability (one minus survival) under exposure level $x = 1$ is

$$\frac{1}{n} \sum_{i=1}^n W(1, \mathbf{Z}_i; \hat{\alpha}) \hat{U}_i(t) \left(I_{T_i \leq t} - Q_k(1, \mathbf{Z}_i, \hat{\beta}^\dagger, \hat{\gamma}^\dagger) \right) + \frac{1}{n} \sum_{i=1}^n Q_k(1, \mathbf{Z}_i, \hat{\beta}^\dagger, \hat{\gamma}^\dagger),$$

where the term $I_{T_i \leq t}$ is defined to be 0 for individuals censored before time t , but it is otherwise fully observed. Similarly, we can substitute in $x = 0$ and take the difference between them to obtain estimates of differences in survival probabilities. A standard error estimator is suggested in Blanche et al. 2023 and implemented in the R package [logitPCWATE](#). This function is basically a wrapper to that function to be consistent with the other estimators.

Value

A list with estimated survival probabilities in each group, their difference, and estimated standard errors.

References

Blanche, Paul Frédéric, Anders Holt, and Thomas Scheike. "On logistic regression with right censored data, with or without competing risks, and its use for estimating treatment effects." *Lifetime data analysis* 29(2):441-482, 2023.

Examples

```
df <- rotterdam
df$time <- pmin(df$rtime, df$dtime) / 365.25
df$status <- ifelse(df$recur == 1 | df$death == 1, 1, 0)
```

```

df$censor <- 1 - df$status
df$chemo <- factor(df$chemo)
drFit <-
  blanche_survdiff(
    oformula = Event(time, status) ~ chemo + year + age + meno +
      size + factor(grade) + nodes + pgr + er + hormon,
    cformula = ~ chemo + meno + size,
    eformula = chemo ~ year + age + meno + size +
      factor(grade) + nodes + pgr + er + hormon,
    times = c(2.5, 5, 7.5),
    data = df
  )
drFit

```

pseudo_survdiff	<i>Estimate the survival probabilities in two treatment groups at several times</i>
-----------------	---

Description

Doubly robust estimation of the survival probabilities in two exposure groups and their difference using the method of Wang (2018). By default it is assumed that censoring is completely independent, but covariate-dependence can be modeled by either specifying the right side of `cformula` and `cens.method` which will be passed to [cumincglm](#).

Usage

```

pseudo_survdiff(
  oformula,
  cformula = ~1,
  cens.method = NULL,
  eformula,
  times,
  data,
  weights = NULL,
  subset = NULL,
  R = 50,
  parallel = "no",
  cl = NULL,
  ncpus = NULL
)

```

Arguments

<code>oformula</code>	The outcome formula, the left side should be a call to Surv
<code>cformula</code>	The censoring model formula, a one sided formula with no outcome
<code>cens.method</code>	Method to use for the censoring model, passed to cumincglm

eformula	The exposure model formula, which will be fit by logistic regression in a call to glm
times	Vector of times at which to estimate the survival probabilities
data	Data frame in which to find the variables in the model formulas
weights	Vector of case weights
subset	Logical vector
R	Number of bootstrap replicates for standard error
parallel	Parallel processing for bootstrap, see boot
cl	Optional cluster if parallel processing, see boot
ncpus	Number of cpus to use for parallel processing, see boot

Details

As presented in Wang 2018, let the jackknife pseudo observation for the i th subject be

$$\hat{S}^i(t) = n\hat{S}(t) - (n-1)\hat{S}^{-i}(t),$$

where $\hat{S}(t)$ is an estimator of the survival probability at time t using all n subjects, and $\hat{S}^{-i}(t)$ is an estimator of the survival probability at time t leaving the i th subject out of the estimation procedure. The assumption about censoring is implied by the way these pseudo observations are computed, i.e., these could be inverse probability of censoring weighted estimators where the censoring distribution may depend on covariates, nonparametric estimators stratified on covariates, or simply the Kaplan-Meier estimators if censoring is assumed to be independent of covariates. We then use a working outcome model for the conditional survival probability:

$$\log[-\log\{v_i(X_i)\}] = \log[-\log\{p(T > t|X_i, \mathbf{Z}_i)\}] = \beta^* X_i + h(X_i, \mathbf{Z}_i; \gamma^*),$$

where $\log(-\log(x))$ is the link function that coincides with how a proportional hazards model fits survival at this time point. It is of note that, unlike the Cox model, this estimator does not assume anything about the survival function prior to time point t . The parameters in the above model are estimated by solving the generalized estimating equations with $\hat{S}^i(t)$ used as the outcome variable. Let $\hat{v}_i(x) = \exp[-\exp\{\hat{\beta}^* x + h(x, \mathbf{Z}_i; \hat{\gamma}^*)\}]$. A logistic regression model for the propensity of the exposure is assumed and estimated. The doubly robust estimator of the difference in survival probability at time t is

$$\frac{1}{n} \sum_{i=1}^n \left[\frac{X_i \hat{S}^i(t) - (X_i - g(\mathbf{Z}_i, \hat{\alpha})) \hat{v}_i(1)}{g(\mathbf{Z}_i, \hat{\alpha})} - \frac{(1 - X_i) \hat{S}^i(t) + (X_i - g(\mathbf{Z}_i, \hat{\alpha})) \hat{v}_i(0)}{1 - g(\mathbf{Z}_i, \hat{\alpha})} \right].$$

Although Wang proposed a variance estimator, it was found to be biased upward, and they suggest instead using bootstrap variance estimation. Thus, we implement the bootstrap variance estimator.

Value

A list with estimated survival probabilities in each group, their difference, and estimated standard errors.

References

Wang, J. A simple, doubly robust, efficient estimator for survival functions using pseudo observations. *Pharmaceutical Statistics*. 2018; 17: 38–48. <https://doi.org/10.1002/pst.1834>

Examples

```
df <- rotterdam
df$time <- pmin(df$ptime, df$dtime) / 365.25
df$status <- ifelse(df$recur == 1 | df$death == 1, 1, 0)
df$ensor <- 1 - df$status
drFit <-
  pseudo_survdiff(
    oformula = Surv(time, status) ~ chemo + year + age + meno +
      size + factor(grade) + nodes + pgr + er + hormon,
    cformula = ~ chemo + meno + size,
    eformula = chemo ~ year + age + meno + size +
      factor(grade) + nodes + pgr + er + hormon,
    times = c(2.5, 5, 7.5),
    data = df
  )
drFit
```

rotterdam

Rotterdam dataset from survival

Description

Rotterdam dataset from survival

Usage

```
rotterdam
```

Format

A data frame with 2982 rows and 15 variables See [rotterdam](#)

sjovan_survdiff *Estimate the survival curves in two treatment groups*

Description

Doubly robust estimation of the survival curves in two exposure groups and their difference using the method of Sjolander and Vansteelandt (2017) .

Usage

```
sjovan_survdiff(
  oformula = NULL,
  ofunc = "survreg",
  oarg = list(),
  cformula = NULL,
  cfunc = "survreg",
  carg = list(),
  eformula = NULL,
  earg = list(),
  method = "DR",
  times = NULL,
  rel.tol = .Machine$double.eps^0.1,
  jacobian.method = "simple",
  se.type = "none",
  data = NULL,
  weights = NULL,
  subset = NULL,
  R = 50,
  parallel = "no",
  cl = NULL,
  ncpus = NULL
)
```

Arguments

oformula	The outcome formula, the left side should be a call to Surv
ofunc	The model type for the outcome, one of "coxph" or "survreg"
oarg	Arguments passed to ofunc
cformula	The censoring model formula, the left side should be a call to Surv
cfunc	The model type for the censoring model, either "coxph" or "survreg"
carg	Arguments passed to cfunc
eformula	The exposure model formula, which will be fit by logistic regression in a call to glm
earg	Arguments passed to the glm for the exposure model
method	Estimation method, either "IPW" or "DR"

times	Vector of times at which to estimate the survival curves
rel.tol	Convergence tolerance
jacobian.method	Method of computing the jacobian, passed to jacobian
se.type	Method of calculating standard errors, either "none" for no standard errors, "sandwich", or "boot"
data	Data frame in which to find the variables in the model formulas
weights	Vector of case weights
subset	Logical vector
R	Number of bootstrap replicates, used if se.type = "boot"
parallel	Parallel processing for bootstrap, see boot
cl	Optional cluster if parallel processing, see boot
ncpus	Number of cpus to use for parallel processing, see boot

Details

Following Bai et al. 2013 and Sjölander and Vansteelandt 2017, consider the following estimating equation for $S_x(t) = p\{T(x) > t\}$:

$$\sum_{i=1}^n \left[S_x(t) - \frac{I_{X_i=x} I_{\tilde{T}_i > t}}{\bar{g}(\mathbf{Z}_i, \boldsymbol{\alpha}) G_c(t, X_i, \mathbf{Z}_i)} - \frac{I_{X_i=x} - \bar{g}(\mathbf{Z}_i, \boldsymbol{\alpha})}{\bar{g}(\mathbf{Z}_i, \boldsymbol{\alpha})} H(t, \mathbf{Z}_i, X = x) - \frac{I_{X_i=x} - \bar{g}(\mathbf{Z}_i, \boldsymbol{\alpha})}{\bar{g}(\mathbf{Z}_i, \boldsymbol{\alpha})} H(t, \mathbf{Z}_i, X = x) \int_0^t \frac{1}{G_c(u, X_i, \mathbf{Z}_i)} dM_c(u, \mathbf{Z}_i, X_i, \tilde{T}_i, \Delta) \right]$$

where $\bar{g}(\mathbf{Z}_i, \boldsymbol{\alpha}) = g(\mathbf{Z}_i, \boldsymbol{\alpha})^x (1 - g(\mathbf{Z}_i, \boldsymbol{\alpha}))^{(1-x)}$, $H(t, \mathbf{Z}, X)$ is a model for $p\{T > t | \mathbf{Z}, X\}$, and $M_c(t, \mathbf{Z}, X, \tilde{T}, \Delta)$ is the martingale increment for the censoring distribution. The above is an unbiased estimating equation for $S_x(t)$ if either $H(t, \mathbf{Z}, X)$ is correctly specified for both censoring and confounding, i.e., is a correctly specified model for $p\{T(x) > t | \mathbf{Z}, X\}$ or both $G_c(u, X, \mathbf{Z})$ and $g(\mathbf{Z}, \boldsymbol{\alpha})$ are correctly specified for censoring and confounding, respectively. To obtain estimates of the difference in survival probabilities, one must specify models for the unknown functions g , G_c , and H , get estimates of those, plug them into the estimating equations, and solve for $S_x(t)$ under $x \in \{0, 1\}$. In this package, one can use semiparametric Cox models or parametric survival models for the outcome and the censoring distributions, and logistic regression for the propensity score model. Bai et al. 2013 provide an expression for a variance estimator that accounts for the uncertainty due to the estimation of the propensity score g and the censoring distribution G_c .

Value

A list with the estimates survival probabilities in each group, their difference, and standard errors, if requested

References

- Xiaofei Bai, Anastasios A Tsiatis, and Sean-M O'Brien. "Doubly-robust estimators of treatment-specific survival distributions in observational studies with stratified sampling." *Biometrics*, 69(4):830–839, 2013.
- Sjölander, Arvid, and Stijn Vansteelandt. "Doubly robust estimation of attributable fractions in survival analysis." *Statistical methods in medical research* 26(2):948-969, 2017.

Examples

```
df <- rotterdam
df$time <- pmin(df$time, df$dtime) / 365.25
df$status <- ifelse(df$recur == 1 | df$death == 1, 1, 0)
df$censor <- 1 - df$status
drFit <-
  sjovan_survdiff(
    oformula = Surv(time, status) ~ chemo + year + age + meno +
      size + factor(grade) + nodes + pgr + er + hormon,
    ofunc = "survreg",
    cformula = Surv(time, censor) ~ chemo + year + age,
    cfunc = "survreg",
    eformula = chemo ~ year + age + meno + size +
      factor(grade) + nodes + pgr + er + hormon,
    method = "DR",
    times = c(2.5, 5, 7.5),
    se.type = "sandwich",
    data = df
  )
drFit
```

Index

* datasets

rotterdam, 6

blanche_survdiff, 2

boot, 5, 8

cumincglm, 4

Event, 2

glm, 2, 5, 7

jacobian, 8

logitIPCWATE, 3

pseudo_survdiff, 4

rotterdam, 6, 6

sjovan_survdiff, 7

Surv, 4, 7