Package ‘CSTE’

September 7, 2021

Version 1.0

Date 2021-08-24

Type Package

Title Covariate Specific Treatment Effect (CSTE) Curve

Description A uniform statistical inferential tool in making
individualized treatment decisions, which implements the methods of
Ma et al. (2017)<DOI:10.1177/0962280214541724>
and Guo et al. (2021)<DOI:10.1080/01621459.2020.1865167>.
It uses a flexible semiparametric modeling strategy for heterogeneous treatment effect
estimation in high-dimensional settings and can gave valid confidence bands. Based on it,
one can find the subgroups of patients that benefit from each treatment, thereby making
individualized treatment selection.

License GPL (>= 2)

Encoding UTF-8

Imports Rcpp (>= 1.0.4), fda, splines, survival

LinkingTo Rcpp

RoxygenNote 7.1.1

Suggests mvtnorm, sigmoid

NeedsCompilation yes

Author Peng Wu [aut],
Wenjie Hu [aut, cre],
Yuhao Deng [aut],
Xiaohua Zhou [aut]

Maintainer Wenjie Hu <huwenjie@pku.edu.cn>

Repository CRAN

Date/Publication 2021-09-07 09:20:08 UTC

R topics documented:

CSTE-package *Covariate specific treatment effect (CSTE) curve.*

Description

Estimation of optimal individualized treatment rules using a covariate specific treatment effect (CSTE) curve. Suppose we have a binary treatment Z, a p-dimensional covariate X, which may be high-dimensional.

If the outcome Y is a binary variable, we consider the CSTE curve, defined by

$$CSTE(x) = logit(E(Y(1) \mid X = x)) - logit(E(Y(0) \mid X = x)),$$

where $logit(u) = \log(u) - \log(1-u)$, $Y(1)$ and $Y(0)$ denote the potential outcomes if the active treatment and the control treatment are received, respectively. In this case, the CSTE curve represents the difference of the logarithm of odds ratio between treated group and control group, which is a common causal quantity in clinical studies for binary outcome. Moreover, $Y = ZY(1) + (1-Z)Y(0)$. Under the unconfoundedness assumption such that $(Y(0), Y(1)) \perp Z \mid X$, the CSTE curve can be re-expressed as

$$CSTE(x) = logit(E(Y \mid X = x, Z = 1)) - logit(E(Y \mid X = x, Z = 1)).$$

Denoting $\mu(X, Z) = E(Y|X, Z)$, we model the logarithm of odds ration as

$$logit(\mu(X, Z)) = g_1(X\beta_1)Z + g_2(X\beta_2),$$

Then we have

$$CSTE(x) = g_1(x\beta_1).$$

If the outcome T is a failure time with right censoring, we let $T(z)$ be the failure time under the corresponding treatment arm $Z = z$ ($z = 0, 1$), and define the potential conditional hazard rate function as

$$\lambda^{(z)}(t \mid x) = \lim_{\Delta t \to 0} \frac{P(t < T(z) \le t + \Delta t \mid T(z) > t, X = x)}{\Delta t}.$$

In most medical research with a survival outcome, the treatment effect is usually represented by the logarithm of a hazard ratio. Hence, the CSTE curve at a fixed time t is defined as

$$CSTE(x, t) = \log\left\{\frac{\lambda^{(1)}(t \mid x)}{\lambda^{(0)}(t \mid x)}\right\}.$$

We focus on a special case that $CSTE(x,t) = CSTE(x)$ in a randomized clinical trial. This assumption holds under many well-known models, such as Cox's model and varying-coefficient Cox's model. Consequently, we assume the following varying coefficient proportional hazard regression model for event time

$$\lambda(t|X,Z) = \lambda_0(t)\exp(g_1(X\beta_1)Z + g_2(X\beta_2)),$$

which implies that

$$CSTE(x) = g_1(x\beta_1)$$

This package facilitate estimating, making inference and predicting $CSTE(x)$ for binary outcome and time to event outcome with right censoring.

Details

The R package CSTE - version 1.0 can be used for three main tasks:

- `cste_bin`: estimate the CSTE curve for binary outcome.
- `cste_surv`: estimate the CSTE curve for time to event outcome with right censoring.
- `predict_cste_bin`: predict the CSTE curve of new subject with binary outcome.
- `predict_cste_surv`: predict the CSTE curve of new subject for time to event outcome with right censoring.
- `cste_bin_SCB`: compute the simultaneous confidence bands (SCB) of CSTE curve for binary outcome.
- `cste_surv_PCI`: compute the point confidence intervals (PCI) of CSTE curve for time to event outcome with right censoring.

Author(s)

Peng Wu, Wenjie Hu, Yuhao Deng, Xiaohua Zhou.

Maintainer: Wenjie Hu <huwenjie@pku.edu.cn>

References

Guo W., Zhou X. and Ma S. (2021). Estimation of Optimal Individualized Treatment Rules Using a Covariate-Specific Treatment Effect Curve with High-dimensional Covariates, *Journal of the American Statistical Association*, 116(533), 309-321.

Ma Y. and Zhou X. (2017). Treatment selection in a randomized clinical trial via covariate-specific treatment effect curves, *Statistical Methods in Medical Research*, 26(1), 124-141.

`cste_bin` *Estimate the CSTE curve for binary outcome.*

Description

Estimate covariate-specific treatment effect (CSTE) curve. Input data contains covariates X, treatment assignment Z and binary outcome Y. The working model is

$$logit(\mu(X, Z)) = g_1(X\beta_1)Z + g_2(X\beta_2),$$

where $\mu(X, Z) = E(Y|X, Z)$. The model implies that $CSTE(x) = g_1(x\beta_1)$.

Usage

```
cste_bin(
  x,
  y,
  z,
  beta_ini = NULL,
  lam = 0,
  nknots = 1,
  max.iter = 200,
  eps = 0.001
)
```

Arguments

`x`	samples of covariates which is a $n * p$ matrix.
`y`	samples of binary outcome which is a $n * 1$ vector.
`z`	samples of treatment indicator which is a $n * 1$ vector.
`beta_ini`	initial values for $(\beta_1', \beta_2')'$, default value is NULL.
`lam`	value of the lasso penalty parameter λ for β_1 and β_2, default value is 0.
`nknots`	number of knots for the B-spline for estimating g_1 and g_2.
`max.iter`	maximum iteration for the algorithm.
`eps`	numeric scalar ≥ 0, the tolerance for the estimation of β_1 and β_2.

Value

A S3 class of cste, which includes:

- `beta1`: estimate of β_1.
- `beta2`: estimate of β_2.
- `B1`: the B-spline basis for estimating g_1.
- `B2`: the B-spline basis for estimating g_2.

- `delta1`: the coefficient of B-spline for estimating g_1.
- `delta2`: the coefficient for B-spline for estimating g_2.
- `iter`: number of iteration.
- `g1`: the estimate of $g_1(X\beta_1)$.
- `g2`: the estimate of $g_2(X\beta_2)$.

References

Guo W., Zhou X. and Ma S. (2021). Estimation of Optimal Individualized Treatment Rules Using a Covariate-Specific Treatment Effect Curve with High-dimensional Covariates, *Journal of the American Statistical Association*, 116(533), 309-321

See Also

`cste_bin_SCB`, `predict_cste_bin`, `select_cste_bin`

Examples

```
## Quick example for the cste

library(mvtnorm)
library(sigmoid)

# --------  Example 1: p = 20 ---------  #
## generate data
n <- 2000
p <- 20
set.seed(100)

# generate X
sigma <- outer(1:p, 1:p, function(i, j){ 2^(-abs(i-j)) } )
X <- rmvnorm(n, mean = rep(0,p), sigma = sigma)
X <- relu(X + 2) - 2
X <- 2 - relu(2 - X)

# generate Z
Z <- rbinom(n, 1, 0.5)

# generate Y
beta1 <- rep(0, p)
beta1[1:3] <- rep(1/sqrt(3), 3)
beta2 <- rep(0, p)
beta2[1:2] <- c(1, -2)/sqrt(5)
mu1 <- X %*% beta1
mu2 <- X %*% beta2
g1 <- mu1*(1 - mu1)
g2 <- exp(mu2)
prob <- sigmoid(g1*Z + g2)
Y <- rbinom(n, 1, prob)

## estimate the CSTE curve
```

```
fit <- cste_bin(X, Y, Z)

## plot
plot(mu1, g1, cex = 0.5, xlim = c(-2,2), ylim = c(-8, 3),
     xlab = expression(X*beta), ylab = expression(g1(X*beta)))
     ord <- order(mu1)
     points(mu1[ord], fit$g1[ord], col = 'blue', cex = 0.5)

## compute 95% simultaneous confidence band (SCB)
res <- cste_bin_SCB(X, fit, alpha = 0.05)

## plot
plot(res$or_x, res$fit_x, col = 'red',
     type="l", lwd=2, lty = 3, ylim = c(-10,8),
     ylab=expression(g1(X*beta)), xlab = expression(X*beta),
     main="Confidence Band")
lines(res$or_x, res$lower_bound, lwd=2.5, col = 'purple', lty=2)
lines(res$or_x, res$upper_bound, lwd=2.5, col = 'purple', lty=2)
abline(h=0, cex = 0.2, lty = 2)
legend("topleft", legend=c("Estimates", "SCB"),
        lwd=c(2, 2.5), lty=c(3,2), col=c('red', 'purple'))


# --------  Example 2: p = 1 ---------  #

## generate data
set.seed(15)
p <- 1
n <- 2000
X <- runif(n)
Z <- rbinom(n, 1, 0.5)
g1 <- 2 * sin(5*X)
g2 <- exp(X-3) * 2
prob <- sigmoid( Z*g1 + g2)
Y <- rbinom(n, 1, prob)

## estimate the CSTE curve
fit <- cste_bin(X, Y, Z)

## simultaneous confidence band (SCB)
X <- as.matrix(X)
res <- cste_bin_SCB(X, fit)

## plot
plot(res$or_x, res$fit_x, col = 'red', type="l", lwd=2,
     lty = 3, xlim = c(0, 1), ylim = c(-4, 4),
     ylab=expression(g1(X)), xlab = expression(X),
     main="Confidence Band")
lines(res$or_x, res$lower_bound, lwd=2.5, col = 'purple', lty=2)
lines(res$or_x, res$upper_bound, lwd=2.5, col = 'purple', lty=2)
abline(h=0, cex = 0.2)
lines(X[order(X)], g1[order(X)], col = 'blue', lwd = 1.5)
legend("topright", legend=c("Estimates", "SCB",'True CSTE Curve'),
```

```
lwd=c(2, 2.5, 1.5), lty=c(3,2,1), col=c('red', 'purple','blue'))
```

`cste_bin_SCB` *Calculate simultaneous confidence bands of CSTE curve for binary outcome.*

Description

This function calculates simultaneous confidence bands of CSTE curve for binary outcome.

Usage

```
cste_bin_SCB(x, fit, alpha = 0.05)
```

Arguments

`x`	samples of predictor, which is a $m * p$ matrix.
`fit`	a S3 class of cste.
`alpha`	the simultaneous confidence bands are of $1 - \alpha$ confidence level.

Value

A list which includes:

- `or_x`: the ordered value of $X\beta_1$.
- `fit_x`: the fitted value of CSTE curve corresponding to `or_x`.
- `lower_bound`: the lower bound of CSTE's simultaneous confidence band.
- `upper_bound`: the upper bound of CSTE's simultaneous confidence band.

References

Guo W., Zhou X. and Ma S. (2021). Estimation of Optimal Individualized Treatment Rules Using a Covariate-Specific Treatment Effect Curve with High-dimensional Covariates, *Journal of the American Statistical Association*, 116(533), 309-321

See Also

`cste_bin`

`cste_surv` *Estimate the CSTE curve for time to event outcome with right censoring.*

Description

Estimate the CSTE curve for time to event outcome with right censoring. The working model is

$$\lambda(t|X, Z) = \lambda_0(t) \exp(g_1(X\beta_1)Z + g_2(X\beta_2)),$$

which implies that $CSTE(x) = g_1(x\beta_1)$.

Usage

```
cste_surv(
  x,
  y,
  z,
  status = NULL,
  beta_ini = NULL,
  lam = 0,
  nknots = 2,
  max.iter = 200,
  eps = 0.001
)
```

Arguments

`x`	samples of covariates which is a $n * p$ matrix.
`y`	samples of time to event which is a $n * 1$ vector.
`z`	samples of treatment indicator which is a $n * 1$ vector.
`status`	samples of censoring indicator which is a $n * 1$ vector, default value is NULL, indicating no censoring.
`beta_ini`	initial values for $(\beta_1', \beta_2')'$, default value is NULL.
`lam`	value of the lasso penalty parameter λ for β_1 and β_2, default value is 0.
`nknots`	number of knots for the B-spline for estimating g_1 and g_2.
`max.iter`	maximum iteration for the algorithm.
`eps`	numeric scalar ≥ 0, the tolerance for the estimation of β_1 and β_2.

Value

A S3 class of cste, which includes:

- `beta1`: estimate of β_1.
- `beta2`: estimate of β_2.

- B1: the B-spline basis for estimating g_1.
- B2: the B-spline basis for estimating g_2.
- delta1: the coefficient of B-spline for estimating g_1.
- delta2: the coefficient for B-spline for estimating g_2.
- iter: number of iteration.
- g1: the estimate for $g_1(X\beta_1)$.
- g2: the estimate for $g_2(X\beta_2)$.

References

Ma Y. and Zhou X. (2017). Treatment selection in a randomized clinical trial via covariate-specific treatment effect curves, *Statistical Methods in Medical Research*, 26(1), 124-141.

See Also

cste_surv_PCI, predict_cste_surv, select_cste_surv

Examples

```
## Quick example for the cste

## generate data
set.seed(313)
X1 <- runif(200,0,20)
X2 <- runif(200,0,20)
X <- cbind(X1, X2)
Z <- c(rep(1,100),rep(0,100))

beta <- c(0.3,0.4)
lambda <- 0.1 * exp(log(X%*%beta/10+0.6)*Z+(X%*%beta)/20)
Time <- -log(runif(200,0,1))/lambda
C <- runif(200,12,15)
S <- as.numeric(Time <= C)
Time <- Time*S + C*(1-S)
## estimate the CSTE curve
fit <- cste_surv(X, Time, Z, S, nknots=3, max.iter=200)
## pointwise confidence interval (PCI)
res <- cste_surv_PCI(fit)

## plot
plot(res$or_x, res$fit_x, col = 'red', type = "l",
     xlim=c(5,12), ylim=c(-1,2), lwd = 2,
     ylab = "CSTE", xlab = "X * beta1",
     main ="Pointwise Confidence interval")
lines(res$or_x, res$lower_bound, lwd = 3, col = 'purple', lty = 2)
lines(res$or_x, res$upper_bound, lwd = 3, col = 'purple', lty = 2)
abline(h = 0, lty = 2, cex = 0.2)
legend("topleft", legend = c("Estimates", "PCI"),
       lwd = c(2,3), lty = c(1, 2), col = c('red','purple'))
```

```
## True CSTE curve
xb <- sort(as.numeric(X%*%beta))
lines(xb, log(xb/10+0.6), type='l', col='blue')
```

`cste_surv_PCI` *Calculate pointwise confidence intervals of CSTE curve for time to event outcome with right censoring.*

Description

This function calculates pointwise confidence intervals of CSTE curve for time to event outcome with right censoring.

Usage

```
cste_surv_PCI(fit, alpha = 0.05)
```

Arguments

`fit`	a S3 class of cste.
`alpha`	the pointwise confidence intervals are of $1 - \alpha$ confidence level.

Value

A list which includes:

- `or_x`: the ordered value of $X\beta_1$.
- `fit_x`: the fitted value of CSTE curve corresponding to `or_x`.
- `lower_bound`: the lower bound of CSTE's pointwise confidence intervals.
- `upper_bound`: the upper bound of CSTE's pointwise confidence intervals.

References

Ma Y. and Zhou X. (2017). Treatment selection in a randomized clinical trial via covariate-specific treatment effect curves, *Statistical Methods in Medical Research*, 26(1), 124-141.

See Also

`cste_surv`

penC *Solve the penalized logistic regression.*

Description

Solve the penalized logistic regression.

Usage

```
penC(x, y, off, beta, lam, pen)
```

Arguments

x	samples of covariates which is a $n * p$ matrix.
y	samples of binary outcome which is a $n * 1$ vector.
off	offset in logistic regression.
beta	initial estimates.
lam	value of the lasso penalty parameter λ for β_1 and β_2.
pen	1: MCP estimator; 2: SCAD estimator.

Value

A numeric vector, estimate of beta

predict_cste_bin *Predict the CSTE curve of new data for binary outcome.*

Description

Predict the CSTE curve of new data for binary outcome.

Usage

```
predict_cste_bin(obj, newx)
```

Arguments

obj	a S3 class of cste.
newx	samples of covariates which is a $m * p$ matrix.

Value

A S3 class of cste which includes

- g1: predicted $g_1(X\beta_1)$.
- g2: predicted $g_2(X\beta_2)$.
- B1: the B-spline basis for estimating g_1.
- B2: the B-spline basis for estimating g_2.

References

Guo W., Zhou X. and Ma S. (2021). Estimation of Optimal Individualized Treatment Rules Using a Covariate-Specific Treatment Effect Curve with High-dimensional Covariates, *Journal of the American Statistical Association*, 116(533), 309-321

See Also

cste_bin

---

predict_cste_surv — *Predict the CSTE curve of new data for time to event outcome with right censoring.*

Description

Predict the CSTE curve of new data for time to event outcome with right censoring.

Usage

```
predict_cste_surv(obj, newx, alpha = 0.05)
```

Arguments

obj	a S3 class of cste.
newx	samples of covariates which is a $m * p$ matrix.
alpha	(1-α) confidence level.

Value

A S3 class of cste, which includes

- g1: predicted $g_1(X\beta_1)$.
- lower_bound: the lower bound of CSTE's pointwise confidence intervals.
- upper_bound: the upper bound of CSTE's pointwise confidence intervals.

References

Ma Y. and Zhou X. (2017). Treatment selection in a randomized clinical trial via covariate-specific treatment effect curves, *Statistical Methods in Medical Research*, 26(1), 124-141.

See Also

cste_surv

---

select_cste_bin — *Select the optimal tuning parameters in CSTE estimation for binary outcome.*

Description

select lasso penalty parameter λ for β_1 and β_2 in CSTE estimation.

Usage

```
select_cste_bin(
  x,
  y,
  z,
  lam_seq,
  beta_ini = NULL,
  nknots = 1,
  max.iter = 2000,
  eps = 0.001
)
```

Arguments

x	samples of covariates which is a $n * p$ matrix.
y	samples of binary outcome which is a $n * 1$ vector.
z	samples of treatment indicator which is a $n * 1$ vector.
lam_seq	a sequence for the choice of λ.
beta_ini	initial values for $(\beta_1', \beta_2')'$, default value is NULL.
nknots	number of knots for the B-spline for estimating g_1 and g_2.
max.iter	maximum iteration for the algorithm.
eps	numeric scalar ≥ 0, the tolerance for the estimation of β_1 and β_2.

Value

A list which includes

- `optimal`: optimal cste within the given the sequence of λ.
- `bic`: BIC for the sequence of λ.
- `lam_seq`: the sequence of λ that is used.

References

Guo W., Zhou X. and Ma S. (2021). Estimation of Optimal Individualized Treatment Rules Using a Covariate-Specific Treatment Effect Curve with High-dimensional Covariates, *Journal of the American Statistical Association*, 116(533), 309-321

See Also

cste_bin

`select_cste_surv` *Select the optimal tuning parameters in CSTE estimation for time to event outcome with right censoring.*

Description

select lasso penalty parameter λ for β_1 and β_2 in CSTE estimation.

Usage

```
select_cste_surv(
  x,
  y,
  z,
  status = NULL,
  lam_seq,
  beta_ini = NULL,
  nknots = 1,
  max.iter = 2000,
  eps = 0.001
)
```

Arguments

x	samples of covariates which is a $n * p$ matrix.
y	samples of binary outcome which is a $n * 1$ vector.
z	samples of treatment indicator which is a $n * 1$ vector.
status	samples of censoring indicator which is a $n * 1$ vector. Default value is NULL, indicating no censoring.
lam_seq	a sequence for the choice of λ.
beta_ini	initial values for $(\beta_1', \beta_2')'$, default value is NULL.
nknots	number of knots for the B-spline for estimating g_1 and g_2.
max.iter	maximum iteration for the algorithm.
eps	numeric scalar ≥ 0, the tolerance for the estimation of β_1 and β_2.

Value

A list which includes

- `optimal`: optimal cste within the given the sequence of λ.
- `bic`: BIC for the sequence of λ.
- `lam_seq`: the sequence of λ that is used.

References

Ma Y. and Zhou X. (2017). Treatment selection in a randomized clinical trial via covariate-specific treatment effect curves, *Statistical Methods in Medical Research*, 26(1), 124-141.

See Also

`cste_surv`

Index