R documentation

of all in '.'

September 22, 2021

R topics documented:

`smle` *Performs efficient semiparametric estimation for regression models with univariate outcomes under general two-phase designs.*

Description

Performs efficient semiparametric estimation for regression models with univariate outcomes under general two-phase designs.

Usage

```
smle(Y = NULL, Delta = NULL, X = NULL, Z = NULL, Bspline_Z = NULL,
  data = NULL, hn_scale = 1, MAX_ITER = 2000, TOL = 1e-04,
  noSE = FALSE, model = "linear", verbose = FALSE)
```

Arguments

`Y`	Specifies the column of the continuous, binary (0 or 1), or censored outcome. Subjects with missing values of `Y` are omitted from the analysis. This argument is required.
`Delta`	Specifies the column of the event indicator. This argument is required when performing proportional hazards regression.
`X`	Specifies the columns of the expensive covariates. Subjects with missing values of `X` are considered as those not selected in the second phase. This argument is required.
`Z`	Specifies the columns of the inexpensive covariates. Subjects with missing values of `Z` are omitted from the analysis. This argument is optional.

Bspline_Z	Specifies the columns of the B-spline basis. Subjects with missing values of Bspline_Z are omitted from the analysis. This argument is not needed when X is independent of Z.
data	Specifies the name of the dataset. This argument is required.
hn_scale	Specifies the scale of the perturbation constant in the variance estimation. For example, if hn_scale = 0.5, then the perturbation constant is $0.5n^{-1/2}$, where n is the first-phase sample size. The default value is 1. This argument is optional. It is not needed when there is no Z.
MAX_ITER	Specifies the maximum number of iterations in the EM algorithm. The default number is 2000. This argument is optional.
TOL	Specifies the convergence criterion in the EM algorithm. The default value is 1E-4. This argument is optional.
noSE	If TRUE, then the variances of the parameter estimators will not be estimated. The default value is FALSE. This argument is optional.
model	Specifies the model. Possible values are "linear", "logistic", and "coxph". The default value is "linear".
verbose	If TRUE, then show details of the analysis. The default value is FALSE.

Value

coefficients	Stores the analysis results.
converge	In parameter estimation, if the EM algorithm converges, then converge = TRUE. Otherwise, converge = FALSE.
converge2	In variance estimation, if the EM algorithm converges, then converge2 = TRUE. Otherwise, converge2 = FALSE.

References

Tao, R., Zeng, D., and Lin, D. Y. (2017). "Efficient Semiparametric Inference Under Two-Phase Sampling, with Applications to Genetic Association Studies", Journal of the American Statistical Association, 112: 1468-1476.

---

smle_lmm *Performs efficient semiparametric estimation for linear mixed models with longitudinal continuous outcomes under general two-phase designs.*

Description

Performs efficient semiparametric estimation for linear mixed models with longitudinal continuous outcomes under general two-phase designs.

Usage

```
smle_lmm(Y = NULL, Time = NULL, ID = NULL, X = NULL, Z = NULL,
  ZT = FALSE, Bspline_Z = NULL, data = NULL, hn_scale = 1,
  MAX_ITER = 2000, TOL = 1e-04, noSE = FALSE, verbose = FALSE,
  coef_initial = NULL, vc_initial = NULL)
```

Arguments

Y	Specifies the column of the longitudinal continuous outcome. Observations with missing values of Y are omitted from the analysis. This argument is required.
Time	Specifies the column of the time variable. Observations with missing values of Time are omitted from the analysis. This argument is required.
ID	Specifies the column of the subject ID. Subjects with missing values of ID are omitted from the analysis. This argument is required.
X	Specifies the columns of the expensive covariates. X should be time-invariant. Subjects with missing values of X are considered as those not selected in the second phase. This argument is required.
Z	Specifies the columns of the inexpensive covariates. Z should be time-invariant. Subjects with missing values of Z are omitted from the analysis. This argument is optional.
ZT	If TRUE, then include the interaction term between Z and Time as fixed effects. The default value is FALSE.
Bspline_Z	Specifies the columns of the B-spline basis. Subjects with missing values of Bspline_Z are omitted from the analysis. This argument is not needed when X is independent of Z.
data	Specifies the name of the dataset. This argument is required.
hn_scale	Specifies the scale of the perturbation constant in the variance estimation. For example, if hn_scale = 0.5, then the perturbation constant is $0.5n^{-1/2}$, where n is the number of observations in the first-phase. The default value is 1. This argument is optional. It is not needed when there is no Z.
MAX_ITER	Specifies the maximum number of iterations in the EM algorithm. The default number is 2000. This argument is optional.
TOL	Specifies the convergence criterion in the EM algorithm. The default value is 1E-4. This argument is optional.
noSE	If TRUE, then the variances of the parameter estimators will not be estimated. The default value is FALSE. This argument is optional.
verbose	If TRUE, then show details of the analysis. The default value is FALSE.
coef_initial	Specifies the initial values of the fixed effects. coef_initial should be a numerical vector. The elements of coef_initial should follow the order of expensive covariates, intercept, inexpensive covariates (if any), time, and interactions between expensive covariates and time. This argument is optional. If this argument is not specified, then the maximum likelihood estimates from the null model without the expensive covariates will be used as initial values.
vc_initial	Specifies the initial values of the variance component parameters. vc_initial should be a numerical vector of length 4. The elements of vc_initial should follow the order of variance of random intercept, covariance between random intercept and random slope of time, variance of random slope of time, and variance of the residual error term. This argument is optional. If this argument is not specified, then the maximum likelihood estimates from the null model without the expensive covariates will be used as initial values.

Value

coefficients	Stores the fixed effects estimates.
vc	Stores the variance component estimates.

converge	In parameter estimation, if the EM algorithm converges, then converge = TRUE. Otherwise, converge = FALSE.
converge2	In variance estimation, if the EM algorithm converges, then converge2 = TRUE. Otherwise, converge2 = FALSE.

---

smle_MEXY *Performs efficient semiparametric estimation for general two-phase measurement error models when there are errors in both the outcome and covariates.*

Description

Performs efficient semiparametric estimation for general two-phase measurement error models when there are errors in both the outcome and covariates.

Usage

```
smle_MEXY(Y_tilde = NULL, Y = NULL, X_tilde = NULL, X = NULL,
  Z = NULL, data = NULL, MAX_ITER = 2000, TOL = 1e-04, noSE = FALSE,
  verbose = FALSE)
```

Arguments

Y_tilde	Specifies the column of the error-prone outcome that is continuous. Subjects with missing values of Y_tilde are omitted from the analysis. This argument is required.
Y	Specifies the column that stores the validated value of Y_tilde in the second phase. Subjects with missing values of Y are considered as those not selected in the second phase. This argument is required.
X_tilde	Specifies the columns of the error-prone covariates. Subjects with missing values of X_tilde are omitted from the analysis. This argument is required.
X	Specifies the columns that store the validated values of X_tilde in the second phase. Subjects with missing values of X are considered as those not selected in the second phase. This argument is required.
Z	Specifies the columns of the accurately measured covariates. Subjects with missing values of Z are omitted from the analysis. This argument is optional.
data	Specifies the name of the dataset. This argument is required.
MAX_ITER	Specifies the maximum number of iterations in the EM algorithm. The default number is 2000. This argument is optional.
TOL	Specifies the convergence criterion in the EM algorithm. The default value is 1E-4. This argument is optional.
noSE	If TRUE, then the variances of the parameter estimators will not be estimated. The default value is FALSE. This argument is optional.
verbose	If TRUE, then show details of the analysis. The default value is FALSE.

Value

`coefficients`	Stores the analysis results.
`converge`	In parameter estimation, if the EM algorithm converges, then `converge = TRUE`. Otherwise, `converge = FALSE`.
`converge2`	In variance estimation, if the EM algorithm converges, then `converge2 = TRUE`. Otherwise, `converge2 = FALSE`.

Index