

R documentation

of all in ‘.’

September 22, 2021

R topics documented:

logreg2ph	1
observed_data_loglik	3
pl_theta	4
profile_out	5

Index	8
--------------	----------

logreg2ph	<i>Sieve maximum likelihood estimator (SMLE) for two-phase logistic regression problems This function returns the sieve maximum likelihood estimators (SMLE) for the logistic regression model from Lotspeich et al. (2021)</i>
-----------	---

Description

Sieve maximum likelihood estimator (SMLE) for two-phase logistic regression problems This function returns the sieve maximum likelihood estimators (SMLE) for the logistic regression model from Lotspeich et al. (2021)

Usage

```
logreg2ph(  
  Y_unval = NULL,  
  Y_val = NULL,  
  X_unval = NULL,  
  X_val = NULL,  
  C = NULL,  
  Validated = NULL,  
  Bspline = NULL,  
  data,  
  theta_pred = NULL,  
  gamma_pred = NULL,  
  initial_lr_params = "Zero",  
  h_N_scale = 1,  
)
```

```

    noSE = FALSE,
    TOL = 1e-04,
    MAX_ITER = 1000
  )

```

Arguments

Y_unval	Column names with the unvalidated outcome. If Y_unval is null, the outcome is assumed to be error-free.
Y_val	Column names with the validated outcome.
X_unval	Column name(s) with the unvalidated predictors. If X_unval and X_val are null, all predictors are assumed to be error-free.
X_val	Column name(s) with the validated predictors. If X_unval and X_val are null, all predictors are assumed to be error-free.
C	(Optional) Column name(s) with additional error-free covariates.
Validated	Column name with the validation indicator. The validation indicator can be defined as Validated = 1 or TRUE if the subject was validated and Validated = 0 or FALSE otherwise.
Bspline	Vector of column names containing the B-spline basis functions.
data	A dataframe with one row per subject containing columns: Y_unval, Y_val, X_unval, X_val, C, Validated, and Bspline.
theta_pred	Vector of columns in data that pertain to the predictors in the analysis model.
gamma_pred	Vector of columns in data that pertain to the predictors in the outcome error model.
initial_lr_params	Initial values for parametric model parameters. Choices include (1) "Zero" (non-informative starting values) or (2) "Complete-data" (estimated based on validated subjects only)
h_N_scale	Size of the perturbation used in estimating the standard errors via profile likelihood. If none is supplied, default is h_N_scale = 1.
noSE	Indicator for whether standard errors are desired. Defaults to noSE = FALSE.
TOL	Tolerance between iterations in the EM algorithm used to define convergence.
MAX_ITER	Maximum number of iterations allowed in the EM algorithm.

Value

Coefficients	dataframe with final coefficient and standard error estimates (where applicable).
converged	indicator of EM algorithm convergence for parameter estimates.
se_converged	indicator of standard error estimate convergence.
converged_msg	(where applicable) description of non-convergence.
iterations	number of iterations completed by EM algorithm to find parameter estimates.

observed_data_loglik *Observed-data log-likelihood*

Description

This function returns the value of the observed-data log-likelihood (equation (2) in Lotspeich et al. (2021)) for a given dataset and parameter values theta, gamma, and p.

Usage

```
observed_data_loglik(
  N,
  n,
  Y_unval = NULL,
  Y_val = NULL,
  X_unval = NULL,
  X_val = NULL,
  C = NULL,
  Bspline = NULL,
  comp_dat_all,
  theta_pred,
  gamma_pred,
  theta,
  gamma,
  p
)
```

Arguments

N	Phase I sample size
n	Phase II sample size
Y_unval	Column names with the unvalidated outcome. If Y_unval is null, the outcome is assumed to be error-free.
Y_val	Column names with the validated outcome.
X_unval	Column name(s) with the unvalidated predictors. If X_unval and X_val are null, all predictors are assumed to be error-free.
X_val	Column name(s) with the validated predictors. If X_unval and X_val are null, all predictors are assumed to be error-free.
C	(Optional) Column name(s) with additional error-free covariates.
Bspline	Vector of column names containing the B-spline basis functions.
comp_dat_all	Augmented dataset containing rows for each combination of unvalidated subjects' data with values from Phase II (a matrix)
theta_pred	Vector of columns in comp_dat_all that pertain to the predictors in the analysis model.
gamma_pred	Vector of columns in comp_dat_all that pertain to the predictors in the outcome error model.
theta	Parameters for the analysis model (a column vector)

gamma	Parameters for the outcome error model (a column vector)
p	B-spline coefficients for the approximated covariate error model (a matrix)
Validated	Column name with the validation indicator. The validation indicator can be defined as Validated = 1 or TRUE if the subject was validated and Validated = 0 or FALSE otherwise.

Value

Scalar value of the function

pl_theta	<i>Profile likelihood for theta, the analysis model parameters</i>
----------	--

Description

This function returns the value of the profile log-likelihood for parameters theta of the analysis model $P(Y|X,C)$ after perturbing element k of theta by some small amount h_N .

Usage

```
pl_theta(
  k,
  theta,
  h_N,
  n,
  N,
  Y_unval,
  Y_val,
  X_unval,
  X_val,
  C,
  Bspline,
  comp_dat_all,
  theta_pred,
  gamma_pred,
  gamma0 = NULL,
  p0 = NULL,
  p_val_num = NULL,
  TOL,
  MAX_ITER
)
```

Arguments

k	A numeric index between 1 and the dimension of theta for the element of theta to be perturbed
theta	Parameters for the analysis model (a column vector) at convergence, resulting from the EM algorithm
h_N	Size of the small perturbation in $\theta[k]$, by default chosen to be $h_N = N^{-1/2}$

n	Phase II sample size
N	Phase I sample size
Y_unval	Column names with the unvalidated outcome. If Y_unval is null, the outcome is assumed to be error-free.
Y_val	Column names with the validated outcome.
X_unval	Column name(s) with the unvalidated predictors. If X_unval and X_val are null, all predictors are assumed to be error-free.
X_val	Column name(s) with the validated predictors. If X_unval and X_val are null, all predictors are assumed to be error-free.
C	(Optional) Column name(s) with additional error-free covariates.
Bspline	Vector of column names containing the B-spline basis functions.
comp_dat_all	Augmented dataset containing rows for each combination of unvalidated subjects' data with values from Phase II (a matrix)
theta_pred	Vector of columns in comp_dat_all that pertain to the predictors in the analysis model.
gamma_pred	Vector of columns in comp_dat_all that pertain to the predictors in the outcome error model.
gamma0	Starting values for gamma, the parameters for the outcome error model (a column vector)
p0	Starting values for p, the B-spline coefficients for the approximated covariate error model (a matrix)
p_val_num	Contributions of validated subjects to the numerator for p, which are fixed (a matrix)
TOL	Tolerance between iterations in the EM algorithm used to define convergence.
MAX_ITER	Maximum number of iterations allowed in the EM algorithm.
Validated	Column name with the validation indicator. The validation indicator can be defined as Validated = 1 or TRUE if the subject was validated and Validated = 0 or FALSE otherwise.

Value

Profile likelihood for theta after perturbing element k by h_N.

profile_out	<i>Profiles out nuisance parameters from the observed-data log-likelihood for a given value of theta</i>
-------------	--

Description

For a given vector theta to parameterize $P(Y|X,C)$, this function repeats the EM algorithm to find the values of gamma and p at convergence. The resulting parameters are used to find the profile log-likelihood for theta by plugging them into the observed-data log-likelihood. This function is used by pl_theta().

Usage

```

profile_out(
  theta,
  n,
  N,
  Y_unval = NULL,
  Y_val = NULL,
  X_unval = NULL,
  X_val = NULL,
  C = NULL,
  Bspline = NULL,
  comp_dat_all,
  theta_pred,
  gamma_pred,
  gamma0,
  p0,
  p_val_num,
  TOL,
  MAX_ITER
)

```

Arguments

theta	Parameters for the analysis model (a column vector)
n	Phase II sample size
N	Phase I sample size
Y_unval	Column names with the unvalidated outcome. If Y_unval is null, the outcome is assumed to be error-free.
Y_val	Column names with the validated outcome.
X_unval	Column name(s) with the unvalidated predictors. If X_unval and X_val are null, all predictors are assumed to be error-free.
X_val	Column name(s) with the validated predictors. If X_unval and X_val are null, all predictors are assumed to be error-free.
C	(Optional) Column name(s) with additional error-free covariates.
Bspline	Vector of column names containing the B-spline basis functions.
comp_dat_all	Augmented dataset containing rows for each combination of unvalidated subjects' data with values from Phase II (a matrix)
theta_pred	Vector of columns in comp_dat_all that pertain to the predictors in the analysis model.
gamma_pred	Vector of columns in comp_dat_all that pertain to the predictors in the outcome error model.
gamma0	Starting values for gamma, the parameters for the outcome error model (a column vector)
p0	Starting values for p, the B-spline coefficients for the approximated covariate error model (a matrix)
p_val_num	Contributions of validated subjects to the numerator for p, which are fixed (a matrix)
TOL	Tolerance between iterations in the EM algorithm used to define convergence.

MAX_ITER	Maximum number of iterations allowed in the EM algorithm.
Validated	Column name with the validation indicator. The validation indicator can be defined as Validated = 1 or TRUE if the subject was validated and Validated = 0 or FALSE otherwise.

Value

Profile likelihood for theta: the value of the observed-data log-likelihood after profiling out other parameters.

Index

logreg2ph, [1](#)

observed_data_loglik, [3](#)

pl_theta, [4](#)

profile_out, [5](#)