Package 'misclassifyr'

September 18, 2024

Title Estimation and Inference for Models of Misclassification.

Version 0.1.4

Description This package provides tools for estimation and inference of simple misclassification models, as described in Mattheis (2024).

License `use_mit_license()`

Encoding UTF-8

Roxygen list(markdown = TRUE)

RoxygenNote 7.3.1

Imports Rcpp,
dplyr,
ggplot2,
parallel,
pbapply,
numDeriv,
pracma,
Matrix

Suggests knitr,
rmarkdown,
testthat (>= 3.0.0)

Config/testthat/edition 3

VignetteBuilder knitr

R topics documented:

`log_prior_Delta_NP` *Evaluates the log of the prior of Delta at model_to_Delta_NP(psi).*

Description

Evaluates the log of the prior of Delta at model_to_Delta_NP(psi).

Usage

```
log_prior_Delta_NP(psi)
```

Arguments

`psi`	A numeric vector parameterizing `Delta` through `model_to_Delta_NP`.

Value

A numeric value equal to the log of the flat prior of `Delta` at `psi`, re-scaled for the logit transform.

`log_prior_Delta_NP_ind` *Evaluates the log of the prior of Delta at model_to_Delta_NP_ind of psi.*

Description

Evaluates the log of the prior of Δ at model_to_Delta_NP_ind(psi).

Usage

```
log_prior_Delta_NP_ind(psi)
```

Arguments

`psi`	A numeric vector parameterizing `Delta` through `model_to_Delta_NP_ind`.

Value

A numeric value equal to the log of the flat prior of `Delta` at `psi`, re-scaled for the logit transform.

`log_prior_Delta_RL_ind` *Evaluates the log of the prior of Delta at model_to_Delta_RL_ind(psi).*

Description

Evaluates the log of the prior of Delta at model_to_Delta_RL_ind(psi).

Usage

```
log_prior_Delta_RL_ind(psi)
```

Arguments

`psi`	A numeric vector parameterizing `Delta` through `model_to_Delta_RL_ind`.

Value

A numeric value equal to the log of the flat prior of `Delta` at `psi`, re-scaled for the logit transform.

`log_prior_Pi_NP` *Evaluates the log of the prior of Pi at model_to_Pi_NP(phi).*

Description

Evaluates the log of the prior of Pi at model_to_Pi_NP(phi).

Usage

```
log_prior_Pi_NP(phi)
```

Arguments

`phi`	A numeric vector parameterizing `Pi` through `model_to_Pi`.

Value

A numeric value equal to the log of the flat prior of `Pi` at `phi`, re-scaled for the logit transform.

misclassifyr *misclassifyr*

Description

This function provides a menu of options for estimation and inference of misclassification models in which the analyst has access to two noisy measures, Y1 and Y2 of a latent outcome Y*, a correctly measured covariate X, and discrete controls W.

Usage

```
misclassifyr(
  tab,
  J,
  K,
  X_names,
  Y1_names,
  Y2_names,
  W_names = NA,
  model_to_Pi = model_to_Pi_NP,
  model_to_Delta = model_to_Delta_NP_ind,
  makeplots = T,
  phi_0 = NA,
  psi_0 = NA,
  misclassification_size = 0.2,
  estimate_beta = F,
  X_vals = NA,
  Y_vals = NA,
  X_col_name = "X",
  Y_col_name = "Y",
  mle = T,
  optim_tol = 1e-08,
  optim_maxit = 1e+05,
  check_stability = F,
  stability_sd = 0.1,
  bayesian = F,
  log_prior_Pi = log_prior_Pi_NP,
  log_prior_Delta = log_prior_Delta_NP_ind,
  n_mcmc_draws = 10000,
  n_burnin = 5000,
  thinning_rate = 1,
  gibbs_proposal_sd = 0.1,
  cores = 1
)
```

Arguments

tab	A dataframe or a list of dataframes containing tabulated data or a list of tabulated data split by controls. The columns should have names Y1, Y2, X, and n where n is a non-negative numeric vector corresponding to the counts of Y1,Y2, and X. The rows should be ordered according to order(Y2,Y1,X).

J	An integer or list corresponding to the number of unique values of Y1 and Y2.
K	An integer or list corresponding to the number of unique values of X.
X_names	A character vector or list corresponding to the values of the regressor X.
Y1_names	A character vector or list corresponding to the values of the outcome Y1.
Y2_names	A character vector or list corresponding to the values of the instrument Y2.
W_names	A character vector corresponding to the values of the control W in each cell.
model_to_Pi	A function or list of functions mapping the parameters of a model for the joint distribution to the joint distribution \eqn{\Pi}.
model_to_Delta	A function or list of functions mapping the parameters of a model to the conditional distribution Y1, Y2 | Y*, \eqn{\Delta}.
makeplots	A logical value for whether to make trace plots and plots of Π and $Delta$. Defaults to TRUE.
phi_0	A numeric vector or list of numeric vectors providing the starting location for optimization for the argument to model_to_Pi.
psi_0	A numeric vector or list of numeric vectors providing the starting location for optimization for the argument to model_to_Delta.
misclassification_size	A numeric value between zero and 1/2 representing a guess for the average share of misclassified values in Y. The initial value for psi_0 will have a diagonal of 1-misclassification_size if psi_0 is not otherwise specified.
estimate_beta	A logical value indicating whether to regress Y on X.
X_vals	A numeric vector or list of numeric vectors providing the values of X associated with the columns of Pi.
Y_vals	A numeric vector or list of numeric vectors providing the values of Y associated with the rows of Pi.
X_col_name	A character vector corresponding to the variable of the regressor X, used only for plots.
Y_col_name	A character vector corresponding to the variable of the outcome Y, used only for plots.
mle	A logical value indicating whether to estimate Pi and Delta via MLE. Defaults to TRUE.
optim_tol	A numeric value giving the relative tolerance for optimization with the optim.
optim_maxit	An integer giving the maximum number of iterations for optim.
check_stability	A logical value indicating whether to perform a more rigorous stability test for the numerical optimizer.
stability_sd	A numerical value giving the standard deviation of the noise added to the initial parameter value for the stability test of the MLE.
bayesian	A logical value indicating whether or not to compute the posterior of values.
log_prior_Pi	A function or list of functions evaluating the log of the prior of Pi at phi (in logs!).
log_prior_Delta	A function or list of functions evaluating the log of the prior of Delta at psi (in logs!).
n_mcmc_draws	An integer corresponding to the length of the MCMC chain.

n_burnin	An integer giving the length of the burn-in period for each MCMC chain, must be shorter than n_mcmc_draws.
thinning_rate	An integer indicating how frequently to record posterior draws from the MCMC chain – e.g. a thinning_rate of 2 records every other draw.
gibbs_proposal_sd	A numeric value giving the standard deviation for the proposal distribution in each Gibbs step.
cores	An integer for the number of CPUs available for parallel processing.
split_eta	An integer or list indicating where to split the vector eta in phi and psi, the arguments to model_to_Pi and model_to_Delta respectively.

Value

A list containing the following components:

- $Pi_hat_MLE: The MLE estimate of the joint distribution of X and Y^*, Π.

---

model_to_Delta_NP *Maps model parameters, psi, to Delta, the fully non-parametric distribution of Y1, Y2 | Y**

Description

Maps model parameters, psi, to Delta, the fully non-parametric distribution of Y1, Y2 | Y*

Usage

```
model_to_Delta_NP(psi)
```

Arguments

psi	A numeric vector of lengthJ^2*(J-1) containing all but the last row of Delta.

Value

A numeric vector corresponding to the JxJ^2 matrix Delta.

---

model_to_Delta_NP_ind *Maps model parameters, psi, to Delta, the distribution of Y1, Y2 | Y* under conditional independence of Y1, Y2 on Y*

Description

Maps model parameters, psi, to Delta, the distribution of Y1, Y2 | Y* under conditional independence of Y1, Y2 on Y

Usage

```
model_to_Delta_NP_ind(psi)
```

Arguments

psi	A numeric vector of length 2xJx(J-1) containing Delta^(1) and Delta^(2).

Value

A numeric vector of length J^3 corresponding to the values of the JxJ^2 matrix Delta.

---

model_to_Delta_RL_ind *Maps model parameters, psi, to Delta, the distribution of Y1, Y2 | Y* under record linkage error structure.*

Description

Maps model parameters, psi, to Delta, the distribution of Y1, Y2 | Y* under record linkage error structure.

Usage

```
model_to_Delta_RL_ind(psi)
```

Arguments

psi	A numeric vector of length 2(J-1)+2J corresponding to the column and row scales of the record linkage.

Value

A numeric vector of length J^3 corresponding to the values of the JxJ^2 matrix Delta.

---

model_to_Pi_NP *Maps model parameters, phi, to the joint distribution of X and Y*, Pi.*

Description

Maps model parameters, phi, to the joint distribution of X and Y*, Pi.

Usage

```
model_to_Pi_NP(phi, J, ...)
```

Arguments

phi	A numeric vector.
J	An integer corresponding to the dimension of Y.
...	Additional, optional arguments.

Value

A numeric vector corresponding to the JxK matrix Pi

Pi_to_beta *Maps the joint distribution, Pi, of X and Y* to a scalar, beta*

Description

Maps the joint distribution, Pi, of X and Y* to a scalar, beta

Usage

```
Pi_to_beta(Pi, X_vals, Y_vals, W_weights)
```

Arguments

Pi	A numeric vector or list of numeric vectors containing the elements of Pi.
X_vals	A numeric vector or a list of numeric vectors representing the scalar values associated with X.
Y_vals	A numeric vector or a list of numeric vectors representing the scalar values associated with Y.
W_weights	A numeric vector representing the sample size of each control cell.

Value

A scalar equal to beta.

prep_misclassification_data *prep_misclassification_data*

Description

This function tabulates data and generates metadata in a format to be used with the misclassifyr() function.

Usage

```
prep_misclassification_data(
  data,
  outcome_1,
  outcome_2,
  regressor,
  outcome_1_bin = NA,
  outcome_2_bin = NA,
  regressor_bin = NA,
  controls = NA,
  weights = NA,
  X_names = NA,
  Y1_names = NA,
  Y2_names = NA,
  record_vals = F,
  round_vals = 2
)
```

Arguments

data	A data.frame containing the outcome variable,
outcome_1	A character string denoting the variable in the dataframe to be used as the first measure of an outcome, Y_1.
outcome_2	A character string denoting the variable in the dataframe to be used as the second measure of an outcome, Y_2.
regressor	A character string denoting the variable in the dataframe to be used as the regressor, X.
outcome_1_bin	A character string denoting the variable in the dataframe to be used to group / bin and average the first measure of an outcome, Y_1.
outcome_2_bin	A character string denoting the variable in the dataframe to be used to group / bin and average the second measure of an outcome, Y_2.
regressor_bin	A character string denoting the variable in the dataframe to be used to group / bin and average the regressor, X.
controls	A character string or vector of character strings denoting the variable/variables to be used as non-parametric controls, W.
weights	A character string denoting a variable containing individual level weights
X_names	A vector of character strings denoting the values of the regressor in the desired order. If NA, as is default, names will be inferred from the data.
Y1_names	A vector of character strings denoting the values of the outcome in the desired order. If NA, as is default, names will be inferred from the data.
Y2_names	A vector of character strings denoting the values of the instrument in the desired order. If NA, as is default, names will be inferred from the data.
record_vals	A logical value indicating whether to record the unique values of the outcomes and the regressor. If record_vals = F, you likely want to order the data by the regressor and outcomes before applying prep_misclassification_data.
round_vals	An integer indicating the precision with which to round the names associated with values of the regressor, outcome, and instrument. Default is 2.

Value

A list of objects including tabulated data to be used in misclassifyr()

---

se_beta_deltamethod *Computes the standard error of beta as function of the joint distribution of X and Y*, Pi*

Description

Computes the standard error of beta as function of the joint distribution of X and Y*, Pi

Usage

```
se_beta_deltamethod(Pi, cov_Pi, X_vals, Y_vals, W_weights)
```

Arguments

`Pi`	A numeric vector or list of numeric vectors containing the elements of Pi.
`cov_Pi`	A numeric vector or a list of numeric vectors representing the covariance of estimates of the elements of Pi.
`X_vals`	A numeric vector or a list of numeric vectors representing the scalar values associated with X.
`Y_vals`	A numeric vector or a list of numeric vectors representing the scalar values associated with Y.
`W_weights`	A numeric vector representing the sample size of each control cell.

Value

A scalar equal to the standard error of beta.

Index