

# Package ‘misclassifyr’

September 4, 2024

**Title** Estimation and Inference for Misclassification Models.

**Version** 0.0.0.9000

**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,  
parallel,  
numDeriv,  
dplyr

**Suggests** testthat (>= 3.0.0)

**Config/testthat/edition** 3

## R topics documented:

log_prior_Delta_NP . . . . .	2
log_prior_Delta_NP_ind . . . . .	2
log_prior_Delta_RL_ind . . . . .	3
log_prior_Pi_NP . . . . .	3
misclassifyr . . . . .	4
model_to_Delta_NP . . . . .	5
model_to_Delta_NP_ind . . . . .	6
model_to_Delta_RL_ind . . . . .	6
model_to_Pi_NP . . . . .	7
Pi_to_beta . . . . .	7
Pi_to_betas . . . . .	8
prep_misclassification_data . . . . .	8

<b>Index</b>	<b>10</b>
--------------	-----------

---

log_prior_Delta_NP	<i>Evaluates the log of the prior of Delta at model_to_Delta_NP(psi).</i>
--------------------	---

---

**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	<i>Evaluates the log of the prior of Delta at model_to_Delta_NP_ind(psi).</i>
------------------------	---

---

**Description**

Evaluates the log of the prior of Delta 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	<i>misclassifyr</i>
--------------	---------------------

---

## 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,
  model_to_Pi = model_to_Pi_NP,
  model_to_Delta = model_to_Delta_NP_ind,
  phi_0 = NA,
  psi_0 = NA,
  X_names = NA,
  Y_names = NA,
  W_names = NA,
  estimate_beta = F,
  estimate_betas = F,
  X_vals = NA,
  Y_vals = NA,
  optim_tol = 1e-08,
  optim_maxit = 1e+05,
  bayesian = F,
  prior_Pi = prior_Pi_NP,
  prior_Delta = prior_Delta_NP_ind,
  n_mcmc_draws = 20000,
  n_burnin = 10000,
  thinning_rate = 2,
  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 be numeric with names Y1, Y2, X, and n where Y1 and Y2 take each value between 1 and J, X takes each value between 1 and K, and
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.
model_to_Pi	A function or list of functions mapping the parameters of a model for the joint distribution to the joint distribution $\pi$ .
model_to_Delta	A function or list of functions mapping the parameters of a model to the conditional distribution $Y1, Y2 \mid Y^*, \Delta$ .

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.
X_names	A character vector or list corresponding to the values of the regressor X.
Y_names	A character vector or list corresponding to the values of the outcome Y.
W_names	A character vector corresponding to the values of the control W in each cell.
estimate_beta	A logical value indicating whether to regress Y on X.
estimate_betas	A logical value indicating whether to regress Y on X within covariate cells.
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.
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.
bayesian	A logical value indicating whether or not to compute the posterior of values.
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.
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!).

**Value**

An object that includes estimates and information from the estimation process

---

model_to_Delta_NP	<i>Maps model parameters, psi, to Delta, the fully non-parametric distribution of Y1, Y2   Y*</i>
-------------------	---

---

**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 length  $J^2(J-1)$  containing all but the last row of Delta.

**Value**

A numeric vector corresponding to the  $J \times J^2$  matrix Delta.

---

model\_to\_Delta\_NP\_ind    *Maps model parameters, psi, to Delta, the distribution of  $Y1, Y2 \mid Y^*$  under conditional independence of  $Y1, Y2$  on  $Y$*

---

**Description**

Maps model parameters, psi, to Delta, the distribution of  $Y1, Y2 \mid Y^*$  under conditional independence of  $Y1, Y2$  on  $Y$

**Usage**

```
model_to_Delta_NP_ind(psi)
```

**Arguments**

psi                      A numeric vector of length  $2 \times J \times (J-1)$  containing  $\Delta^{(1)}$  and  $\Delta^{(2)}$ .

**Value**

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

---

model\_to\_Delta\_RL\_ind    *Maps model parameters, psi, to Delta, the distribution of  $Y1, Y2 \mid Y^*$  under record linkage error structure.*

---

**Description**

Maps model parameters, psi, to Delta, the distribution of  $Y1, Y2 \mid Y^*$  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  $J \times J^2$  matrix Delta.

---

model_to_Pi_NP	<i>Maps model parameters, phi, to the joint distribution of X and Y*, Pi.</i>
----------------	---

---

**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	<i>Maps the joint distribution, Pi, of X and Y* to a scalar, beta</i>
------------	---

---

**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.

---

Pi_to_betas	<i>Maps the joint distribution, <math>P_i</math>, of <math>X</math> and <math>Y^*</math> to a vector, representing beta in each covariate cell</i>
-------------	--

---

### Description

Longer description of what it does...

### Usage

```
Pi_to_betas(Pi, X_vals, Y_vals)
```

### Arguments

Pi	A list of numeric vectors containing the elements of $P_i$ .
X_vals	A list of numeric vectors representing the scalar values associated with $X$ .
Y_vals	A list of numeric vectors representing the scalar values associated with $Y$ .

### Details

some details.

### Value

A scalar equal to beta.

### Examples

```
## Not run:
some example code # Should return something

## End(Not run)
```

---

prep_misclassification_data	<i>prep_misclassification_data</i>
-----------------------------	------------------------------------

---

### 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,  
  controls = NA,  
  weights = NA,  
  record_vals = F  
)
```

**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 a 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
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.

**Value**

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

# Index

log\_prior\_Delta\_NP, [2](#)  
log\_prior\_Delta\_NP\_ind, [2](#)  
log\_prior\_Delta\_RL\_ind, [3](#)  
log\_prior\_Pi\_NP, [3](#)  
  
misclassify, [4](#)  
model\_to\_Delta\_NP, [5](#)  
model\_to\_Delta\_NP\_ind, [6](#)  
model\_to\_Delta\_RL\_ind, [6](#)  
model\_to\_Pi\_NP, [7](#)  
  
Pi\_to\_beta, [7](#)  
Pi\_to\_betas, [8](#)  
prep\_misclassification\_data, [8](#)