

# Package ‘Metatron’

February 19, 2015

**Version** 0.1-1

**Date** 2013-10-16

**Title** Meta-analysis for Classification Data and Correction to Imperfect Reference

**Author** Huiling Huang

**Maintainer** Huiling Huang <huiling.huang23@gmail.com>

**Depends** R (>= 2.15.0), lme4, mpt, Matrix

**Description** This package allows doing meta-analysis for primary studies with classification outcomes in order to evaluate systematically the accuracies of classifiers, namely, the diagnostic tests. It provides functions to fit the bivariate model of Reitsma et al.(2005). Moreover, if the reference employed in the classification process isn't a gold standard, its deficit can be detected and its influence to the underestimation of the diagnostic test's accuracy can be corrected, as described in Botella et al.(2013).

**License** GPL (>= 2)

**NeedsCompilation** no

**Repository** CRAN

**Date/Publication** 2014-10-16 12:13:09

## R topics documented:

AUDIT . . . . .	2
ccp . . . . .	2
correct.trees . . . . .	4
fit.bivar . . . . .	5
imperfect.trees . . . . .	8
MMSE . . . . .	9
perfect.trees . . . . .	10
Schuetz . . . . .	12
<b>Index</b>	<b>13</b>

---

 AUDIT

*Alcohol Use Disorders Identification Test*


---

**Description**

Four primary studies which validated the Alcohol Use Disorders Identification Test by comparing the screening outcomes against the reference of self-report alcohol consumption.

**Usage**

AUDIT

**Format**

A data frame containing 5 variables and 4 observations. The variables are:

<b>study</b>	factor	primary study name
<b>TP</b>	numeric	number of true positives
<b>FN</b>	numeric	number of false negatives
<b>FP</b>	numeric	number of false positives
<b>TN</b>	numeric	number of true negatives

**Source**

Botella et al. 2013, table 2

**References**

Botella, J., Huang, H., Suero, M.(2013). Multinomial tree models for assessing the status of the reference in studies of the accuracy of tools for binary classification. *Frontiers in Psychology*,4:694. [http://www.frontiersin.org/Journal/Abstract.aspx?s=956&name=quantitative\\_psychology\\_and\\_measurement&ART\\_DOI=10.3389/fpsyg.2013.00694](http://www.frontiersin.org/Journal/Abstract.aspx?s=956&name=quantitative_psychology_and_measurement&ART_DOI=10.3389/fpsyg.2013.00694)

---

 ccp

*Anti-cyclic Citrullinated Peptide Antibody*


---

**Description**

Data from Nishimura (2007) of anti-cyclic citrullinated peptide antibody (anti-CCP) which report the four frequencies of the classification outcomes.

**Usage**

ccp

*ccp*

3

**Format**

A data frame containing 7 variables and 37 observations. The variables are:

<b>test</b>	factor	name of the screening test
<b>study_id</b>	factor	primary study name
<b>generation</b>	factor	CCP generation used in the study, can be used as covariate.
<b>TP</b>	numeric	number of true positives
<b>FN</b>	numeric	number of false negatives
<b>FP</b>	numeric	number of false positives
<b>TN</b>	numeric	number of true negatives

### Source

Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy, Chapter 10

### References

Nishimura K, Sugiyama D, Kogata Y, Tsuji G, Nakazawa T, Kawano S, Saigo K, Morinobu A, Koshiba M, Kuntz KM, Kamae I, Kumagai S. Meta-analysis:diagnostic accuracy of anti-cyclic citrullinated peptide antibody and rheumatoid factor for rheumatoid arthritis. *Ann Intern Med* 2007; 146:797-808

---

correct.trees

*Correct the Classification Outcomes by Methods of Botella et al (2013)*

---

### Description

This function corrects the frequencies reported in the primary studies when a imperfect reference was used in the classification process. According to the conjoint parameter estimates of multinomial tree model 2 by Botella et al (2013) given by the `imperfect.trees`, one can obtain the corrected frequencies as if the a perfect standard reference were used. Note that correction should be used only if the model 2 is the optimal one compared to model 1.

### Usage

```
correct.trees(x, TP, FN, TN, FP, study, data)
```

### Arguments

x	an object of class "imperfect.trees".
TP	the true positive counts reported in primary studies.Vector of integers,need to be specied either directly or by referring to a variable in data frame.
FN	the false negative counts reported in primary studies.Vector of integers,need to be specied either directly or by referring to a variable in data frame.
TN	the true negative counts reported in primary studies.Vector of integers,need to be specied either directly or by referring to a variable in data frame.
FP	the false positive counts reported in primary studies.Vector of integers,need to be specied either directly or by referring to a variable in data frame.
study	study names or identities. Character vector, need to be specied either directly or by referring to a variable in data frame.
data	optional data frame that contains the above-mentioned variables.

## Details

The study names and their order used by this function to correct frequencies need to be exactly the same as in the previous step which fits the multinomial tree model 2 assuming an imperfect reference. Otherwise it will give an error indicating the problem.

## Value

A data frame with the 4 corrected frequencies for each primary study, namely, TPnew, FNnew, FPnew, TNnew. If the input data contains other information such as covariates, these will be kept unchanged and well matched with the newly corrected frequencies. This data frame can be further used to fit the bivariate model of Reitsma et al. (2005) by the `fit.bivar` function.

## Author(s)

Huiling Huang <huiling.huang23@gmail.com>

## References

Botella, J., Huang, H., Suero, M.(2013). Multinomial tree models for assessing the status of the reference in studies of the accuracy of tools for binary classification. *Frontiers in Psychology*.4:694. [http://www.frontiersin.org/Journal/Abstract.aspx?s=956&name=quantitative\\_psychology\\_and\\_measurement&ART\\_DOI=10.3389/fpsyg.2013.00694](http://www.frontiersin.org/Journal/Abstract.aspx?s=956&name=quantitative_psychology_and_measurement&ART_DOI=10.3389/fpsyg.2013.00694)

Botella, J., Huang, H., Suero, M. Meta-analysis of the accuracy of tools used for binary classification when the primary studies employ different references. To appear in *Psychological Methods*.

## See Also

[imperfect.trees.](#)

## Examples

```
## data of the screening tool Mini Mental State Examination (MMSE) from Botella et al.(2013)
data(MMSE)
## fit the multinomial tree model 2, imperfect reference
(mmse2<-imperfect.trees(TP=TP, FN=FN, TN=TN, FP=FP, study=study, data=MMSE))
## after comparing to the results of model 1, the model 2 is chosen,
## then comes frequency correction.
correct.trees(mmse2, TP=TP, FN=FN, TN=TN, FP=FP, study=study, data=MMSE)
```

## Description

The function fits the bivariate model of Reitsma et al. (2005) by directly modeling the binomial error and normally distributed random effect structure. We specify the model as a generalized linear mixed model using the `glmer` function in package `lme4`, similar to the computational approach by Macaskill (2004) and Harbord et al. (2007). Thus the normal-normal approximation and the MCMC approach can be avoided. A single covariate can be incorporated into the model, users should choose if it affects both the sensitivity and specificity, only sensitivity or only specificity. Results are nearly identical to those by using the `Proc NLMIXED` in SAS.

## Usage

```
fit.bivar(TP, FN, TN, FP, study, data, mods=NULL, covarying="both", ...)
```

```
## S3 method for class 'fit.bivar'
print(x, ...)
## S3 method for class 'fit.bivar'
summary(object, level=0.95, ...)
```

## Arguments

TP	the true positive counts reported in primary studies. Vector of integers, need to be specified either directly or by referring to a variable in data frame.
FN	the false negative counts reported in primary studies. Vector of integers, need to be specified either directly or by referring to a variable in data frame.
TN	the true negative counts reported in primary studies. Vector of integers, need to be specified either directly or by referring to a variable in data frame.
FP	the false positive counts reported in primary studies. Vector of integers, need to be specified either directly or by referring to a variable in data frame.
study	study names or identities. Vector of characters, need to be specified either directly or by referring to a variable in data frame.
data	optional data frame that contains the above-mentioned variables and possibly other information such as the covariate.
mods	optional argument to include a single study-level covariate in the model. Vector specifying the values of the covariate. Default is <code>NULL</code> .
covarying	options to specify the influence of the covariate when it's present, one of the character strings "both", "only sensitivity", "only specificity" can be selected. Default is "both".
x	an object of class "fit.bivar" (for print).
object	an object of class "fit.bivar" (for summary).
level	confidence interval level (for summary). Default is 0.95
...	further arguments to be passed to or from other functions

## Details

To specify the data, either directly input the TP, FN, FP, TN and study as vectors, or referring the corresponding variable names in a data frame. The argument `study` gives the study names, which means if there are several classification procedures within a study, they must have the same value at this argument, but different values in covariate to distinguish them. To specify the model, one should assign proper values to the `mods` and `covarying` arguments.

## Value

An object of class `fit.bivar`, basically a list with the model specification and conventional model fit results, such as goodness of fit statistics and parameter estimates, etc.

The print functions displays the basic model fit outcomes, including the estimates of the combined logit-transformed sensitivity and specificity, as well as those at different levels of the covariate if there is any; the random effects coefficients shown as variance-covariance matrix of the combined logit-transformed sensitivity and specificity; and the goodness of fit statistics are also given. The print function returns no object.

The summary function shows additional information such as standard errors and covariance matrix of the parameters which can be useful later in drawing the SROC curve.

## Author(s)

Huiling Huang <huiling.huang23@gmail.com>

## References

- Macaskill, P. (2004). Empirical Bayes estimates generated in a hierarchical summary ROC analysis agreed closely with those of a full Bayesian analysis. *Journal of Clinical Epidemiology*, 57, 925-32.
- Harbord, R., Deeks, J., Egger, M., Whiting, P., & Sterne, J. (2007). A unification of models for meta-analysis of diagnostic accuracy studies. *Biostatistics*, 8, 239-251
- Reitsma, J., Glas, A., Rutjes, A., Scholten, R., Bossuyt, P., & Zwinderman, A. (2005). Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *Journal of Clinical Epidemiology*, 58, 982-990.

## See Also

[lmer](#).

## Examples

```
## fit bivariate model without covariate to data from a review(Nishimura 2007)
data(ccp)
(ccp.without<-fit.bivar(TP=TP, FN=FN, TN=TN, FP=FP, study=study_id, data=ccp))
summary(ccp.without)
## fit bivariate model with covariate "generation" to the same data set.
(ccp.generation<-fit.bivar(TP=TP, FN=FN, TN=TN, FP=FP, study=study_id, data=ccp, mods=generation,
                          covarying="both"))
summary(ccp.generation)
##fit bivariate model with covariate "Test" to the data from Schuetz(2010)
##comparing the accuracy of CI and MRI
```

```
data(Schuetz)
(CTvsMRI<-fit.bivar(TP=tp, FN=fn, TN=tn, FP=fp, study=Study_ID, data=Schuetz, mods=Test))
summary(CTvsMRI)
```

---

imperfect.trees

*Fit the Multinomial Tree Model 2 of Botella et al (2013)*


---

## Description

The function fits the multinomial tree model 2 of Botella et al (2013) which detects the possible deficit of the standard reference employed in the classification processes. The model is fitted by the EM algorithm implemented in the package `mpt`. Results should be compared to those of multinomial tree model 1 given by `perfect.trees` before making inference.

## Usage

```
imperfect.trees(TP, FN, TN, FP, study, data, ...)

## S3 method for class 'imperfect.trees'
print(x, ...)
```

## Arguments

TP	the true positive counts reported in primary studies. Vector of integers, need to be specified either directly or by referring to a variable in data frame.
FN	the false negative counts reported in primary studies. Vector of integers, need to be specified either directly or by referring to a variable in data frame.
TN	the true negative counts reported in primary studies. Vector of integers, need to be specified either directly or by referring to a variable in data frame.
FP	the false positive counts reported in primary studies. Vector of integers, need to be specified either directly or by referring to a variable in data frame.
study	study names or identities. Vector of characters, need to be specified either directly or by referring to a variable in data frame.
data	optional data frame that contains the above-mentioned variables.
x	an object of class "imperfect.trees" (for <code>print</code> ).
...	further arguments to be passed to or from other functions

## Details

To specify the data, either directly input the TP, FN, FP, TN and study as vectors, or referring the corresponding variable names in a data frame. The multinomial tree models are then generated automatically according to the study number.



**Value**

An object of class "imperfect.trees", basically a list with the model specification and conventional model fit information, such as goodness of fit statistics and parameter estimates, etc. The print function doesn't return an object.

**Author(s)**

Huiling Huang <huiling.huang23@gmail.com>

**References**

Botella, J., Huang, H., Suero, M.(2013). Multinomial tree models for assessing the status of the reference in studies of the accuracy of tools for binary classification. *Frontiers in Psychology*.4:694. [http://www.frontiersin.org/Journal/Abstract.aspx?s=956&name=quantitative\\_psychology\\_and\\_measurement&ART\\_DOI=10.3389/fpsyg.2013.00694](http://www.frontiersin.org/Journal/Abstract.aspx?s=956&name=quantitative_psychology_and_measurement&ART_DOI=10.3389/fpsyg.2013.00694)

Botella, J., Huang, H., Suero, M. Meta-analysis of the accuracy of tools used for binary classification when the primary studies employ different references. To appear in *Psychological Methods*.

**See Also**

[perfect.trees](#).

**Examples**

```
## load data of the screening outcome of Mini Mental State Examination (MMSE)
data(MMSE)
## fit the multinomial tree model 2, imperfect reference
imperfect.trees(TP=TP, FN=FN, TN=TN, FP=FP, study=study, data=MMSE)
```

---

MMSE

*Mini Mental State Examination*

---

**Description**

Four primary studies which report the accuracy of the Mini Mental State Examination by giving the screening outcomes against the Cambridge Mental Disorders of the Elderly Examination (CAMDEX).

**Usage**

MMSE

**Format**

A vector containing 5 variables and 4 observations. The variables are:

<b>TP</b>	numeric	number of true positives
<b>FN</b>	numeric	number of false negatives
<b>FP</b>	numeric	number of false positives
<b>TN</b>	numeric	number of true negatives
<b>study</b>	factor	primary study name

**Source**

Botella et al. 2013, table 2

**References**

Botella, J., Huang, H., Suero, M.(2013). Multinomial tree models for assessing the status of the reference in studies of the accuracy of tools for binary classification. *Frontiers in Psychology*,4:694. [http://www.frontiersin.org/Journal/Abstract.aspx?s=956&name=quantitative\\_psychology\\_and\\_measurement&ART\\_DOI=10.3389/fpsyg.2013.00694](http://www.frontiersin.org/Journal/Abstract.aspx?s=956&name=quantitative_psychology_and_measurement&ART_DOI=10.3389/fpsyg.2013.00694)

---

perfect.trees

*Fit the Multinomial Tree Model 1 of Botella et al (2013)*

---

**Description**

The function fits the multinomial tree model 1 of Botella et al (2013) which assumes that both the sensitivity and specificity of the standard reference employed in the classification processes are equal to 1. The model is fitted by the EM algorithm implemented in the package mpt. Results should be compared to those of multinomial tree model 2 given by perfect.trees before making inference.

**Usage**

```
perfect.trees(TP, FN, TN, FP, study, data)
```

```
## S3 method for class 'perfect.trees'
print(x, ...)
```

**Arguments**

TP	the true positive counts reported in primary studies. Vector of integers, need to be specified either directly or by referring to a variable in data frame.
FN	the false negative counts reported in primary studies. Vector of integers, need to be specified either directly or by referring to a variable in data frame.
TN	the true negative counts reported in primary studies. Vector of integers, need to be specified either directly or by referring to a variable in data frame.

FP	the false positive counts reported in primary studies. Vector of integers, need to be specified either directly or by referring to a variable in data frame.
study	study names or identities. Vector of characters, need to be specified either directly or by referring to a variable in data frame.
data	optional data frame that contains the above-mentioned variables.
x	an object of class "perfect.trees" (for print).
...	further arguments to be passed to or from other functions

### Details

To specify the data, either directly input the TP, FN, FP, TN and study as vectors, or referring the corresponding variable names in a data frame. The multinomial tree models are then generated automatically according to the study number.

### Value

An object of class "perfect.trees", basically a list with the model specification and conventional model fit information, such as goodness of fit statistics and parameter estimates, etc. The print function does not return an object.

### Author(s)

Huiling Huang <huiling.huang23@gmail.com>

### References

Botella, J., Huang, H., Suero, M. (2013). Multinomial tree models for assessing the status of the reference in studies of the accuracy of tools for binary classification. *Frontiers in Psychology*. 4:694. [http://www.frontiersin.org/Journal/Abstract.aspx?s=956&name=quantitative\\_psychology\\_and\\_measurement&ART\\_DOI=10.3389/fpsyg.2013.00694](http://www.frontiersin.org/Journal/Abstract.aspx?s=956&name=quantitative_psychology_and_measurement&ART_DOI=10.3389/fpsyg.2013.00694)

Botella, J., Huang, H., Suero, M. Meta-analysis of the accuracy of tools used for binary classification when the primary studies employ different references. To appear in *Psychological Methods*.

### See Also

[imperfect.trees](#).

### Examples

```
## load data of the accuracy of the screening tool Alcohol Use Disorders Identification Test (AUDIT)
data(AUDIT)
## fitting the multinomial tree model 1
perfect.trees(TP=TP, FN=FN, TN=TN, FP=FP, study=study, data=AUDIT)
## data of the screening tool Mini Mental State Examination (MMSE)
data(MMSE)
## fit the multinomial tree model 1, assuming imperfect reference
perfect.trees(TP=TP, FN=FN, TN=TN, FP=FP, study=study, data=MMSE)
```

---

 Schuetz

*CT vs. MRI for the Diagnosis of Coronary Artery Disease*


---

**Description**

Data from Schuetz (2010) of the diagnosis of coronary artery disease (CAD), used conventional coronary angiography as the reference standard. Among the total 103 studies, 84 evaluated only the accuracy of CT, 14 evaluated only MRI, and 5 studies evaluated both.

**Usage**

Schuetz

**Format**

A data frame containing 6 variables and 108 observations, including 103 studies. The variables are:

<b>Test</b>	factor	name of the screening test
<b>Study_ID</b>	factor	primary study name
<b>tp</b>	numeric	number of true positives
<b>fp</b>	numeric	number of false positives
<b>fn</b>	numeric	number of false negatives
<b>tn</b>	numeric	number of true negatives

**Source**

Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy, Chapter 10

**References**

Schuetz GM, Zacharopoulou NM, Schlattmann P, Dewey M. Meta-analysis: noninvasive coronary angiography using computed tomography versus magnetic resonance imaging. *Annals of Internal Medicine* 2010;152:167-177.

# Index

\*Topic **bivariate model**

fit.bivar, 5

\*Topic **datasets**

AUDIT, 2

ccp, 2

MMSE, 9

Schuetz, 12

\*Topic **file**

correct.trees, 4

imperfect.trees, 8

perfect.trees, 10

AUDIT, 2

ccp, 2

correct.trees, 4

fit.bivar, 5

imperfect.trees, 5, 8, 11

lmer, 7

MMSE, 9

perfect.trees, 9, 10

print.fit.bivar (fit.bivar), 5

print.imperfect.trees

(imperfect.trees), 8

print.perfect.trees (perfect.trees), 10

Schuetz, 12

summary.fit.bivar (fit.bivar), 5