Package: gIPFrm (via r-universe)

August 30, 2024

Type Package

Title Generalized Iterative Proportional Fitting for Relational Models

Version 3.1

Date 2017-07-21

Author Anna Klimova, Tamas Rudas

Maintainer Anna Klimova <aklimova25@gmail.com>

Description Maximum likelihood estimation under relational models, with or without the overall effect.

License GPL-2

NeedsCompilation no

Date/Publication 2017-07-24 09:49:06 UTC

Repository https://annaklimovauw.r-universe.dev

RemoteUrl https://github.com/cran/gIPFrm

RemoteRef HEAD

RemoteSha ec662590567da2900f2bfc8fbd4059eee7de9915

Contents

`gIPFrm-package` *Generalized Iterative Proportional Fitting for Relational Models*

Description

The package provides an iterative scaling procedure that computes the maximum likelihood estimates of the cell frequencies and of the model parameters under a relational model, with or without the overall effect.

Details

Package:	gIPFrm
Type:	Package
Version:	3.1
Date:	2017-07-21
License:	GPL (>= 2)

The iterative proportional fitting procedure is called by the function `g.ipf`.

Note

Tamas Rudas was supported in part by Grant K-106154 from the Hungarian National Scientific Research Fund (OTKA). The authors wish to thank Juraj Medzihorsky for his help with building this package.

Author(s)

Anna Klimova, Tamas Rudas

Maintainer: Anna Klimova <aklimova25@gmail.com>

References

A.Klimova, T.Rudas, A.Dobra, Relational models for contingency tables. J. Multivariate Anal., 2012, 104, 159–173.

A.Klimova, T.Rudas, Iterative proportional scaling for curved exponential families. Scand. J. Statist., 2015, 42, 832–847.

A. Klimova, Coordinate-Free Exponential Families on Contingency Tables. PhD thesis. Advisers: Tamas Rudas and Thomas Richardson.

A.Agresti, Categorical Data Analysis. Wiley, New York, 1990.

J.Aitchison, S.D.Silvey, Maximum-likelihood estimation procedures and associated tests of significance. J. Roy. Statist. Soc. Ser.B, 1960, 22, 154–171.

G.Kawamura, T.Matsuoka, T.Tajiri, M.Nishida, M.Hayashi, Effectiveness of a sugarcane-fish combination as bait in trapping swimming crabs. Fisheries Research, 1995, 22, 155–160.

Examples

```
### Multiplicative model from Aitchison and Silvey (1960)

A = matrix(c(1, 0, 0, 1, 0, 1, 1,
             0, 1, 0, 1, 1, 0, 1,
             0, 0, 1, 0, 1, 1, 1), byrow=TRUE, nrow=3) ## the model matrix

y = c(46,24,7,15,3,4,1) ## the observed data

g.ipf(A, y, 1e-4, "probabilities", "grid")
g.ipf(A, y, 1e-6, "probabilities", "bisection")
```

`bisection.update` *Search for the adjustment factor corresponding to the MLE using the bisection method*

Description

`bisection.update` computes the segment that is known to contain the adjustment factor corresponding to the MLE and finds this adjustment factor using the bisection method. It is needed only for relational models for probabilities.

Usage

```
bisection.update(ModelMx, ObsTbl, tolerance)
```

Arguments

`ModelMx`	an `I` by `J` model matrix of a relational model. Here `I` is the number of observations and `J` is the number of generating subsets.
`ObsTbl`	a vector of observed cell frequencies of length `I`.
`tolerance`	tolerance used in stopping criteria.

Value

`gamma.tilde`	the adjustment factor under the precision given by `tolerance`.
`model.tilde`	the value returned by `ipf.gamma()` with the adjustment factor `gamma` equal to `gamma.tilde`.

Author(s)

Anna Klimova, Tamas Rudas

References

A. Klimova, Coordinate-Free Exponential Families on Contingency Tables. PhD thesis. Advisers: Tamas Rudas and Thomas Richardson.

D. Bertsekas, Non-Linear Programming. Athena Scientific, 1999.

Examples

```
### Multiplicative model from Aitchison and Silvey (1960)

A = matrix(c(1, 0, 0, 1, 0, 1, 1,
             0, 1, 0, 1, 1, 0, 1,
             0, 0, 1, 0, 1, 1, 1), byrow=TRUE, nrow=3) ## the model matrix

y = c(46,24,7,15,3,4,1) ## the observed data

bisection.update(A, y, 1e-4)



## The model of independence for a 2 by 2 contingency table

A = matrix(c( 1,1,0,0,
              0,0,1,1,
              1,0,1,0,
              0,1,0,1), byrow=TRUE, nrow=4) ## the model matrix

y = c(1,2,3,4)  ## the observed data


bisection.update(A, y, 1e-5)
```

---

g.ipf *Generalized Iterative Proportional Fitting for Relational Models*

Description

g.ipf computes the maximum likelihood estimates of the cell frequencies and of the model parameters under a relational model specified by a model matrix.

Usage

```
g.ipf(ModelMatrix, ObsTable, tol, estimand, adjustment)
```

Arguments

ModelMatrix	an I by J model matrix of a relational model. Here I is the number of observations and J is the number of generating subsets.
ObsTable	a vector of observed cell frequencies of length I.
tol	tolerance used in stopping criteria.
estimand	set to "probabilities" in the case of multinomial sampling; set to "intensities" in the case of Poisson sampling.
adjustment	set to "grid" if a grid is used to update the adjustment factor; set to "bisection" if the bisection method is used to update the adjustment factor; set to "none" if estimand is "intensities".

Value

`model.matrix` the model matrix.

`observed.data` the vector of observed cell frequencies.

`fitted.values` the maximum likelihood estimates of the cell frequencies.

`estimated.total` the sum of the estimated cell frequencies.

`adjustment.for.total` the estimated total divided by the observed total.

`adjustment.for.subsets` the adjustment factor for the subset sums.

`model.parameters` the maximum likelihood estimates for model parameters on the multiplicative scale.

`degrees.of.freedom` the degrees of freedom of the relational model.

`chisq.statistic` Pearson's chi-squared statistic.

`p.value.chisq` the p-value, based on Pearson's chi-squared statistic.

`log.likelihood.ratio.statistic` the log likelihood ratio statistic.

`p.value.log.likelihood.ratio.statistic` the p-value, based on the log likelihood ratio statistic.

`Bregman.statistic` the Bregman statistic.

`p.value.Bregman.statistic` the p-value, based on the Bregman statistic.

Author(s)

Anna Klimova, Tamas Rudas

References

A.Klimova, T.Rudas, A.Dobra, Relational models for contingency tables. J. Multivariate Anal., 2012, 104, 159–173.

A.Klimova, T.Rudas, Iterative proportional scaling for curved exponential families. Scand. J. Statist., 2015, 42, 832–847.

A. Klimova, Coordinate-Free Exponential Families on Contingency Tables. PhD thesis. Advisers: Tamas Rudas and Thomas Richardson.

A.Agresti, Categorical Data Analysis. Wiley, New York, 1990.

J.Aitchison, S.D.Silvey, Maximum-likelihood estimation procedures and associated tests of significance. J. Roy. Statist. Soc. Ser.B, 1960, 22, 154–171.

G.Kawamura, T.Matsuoka, T.Tajiri, M.Nishida, M.Hayashi, Effectiveness of a sugarcane-fish combination as bait in trapping swimming crabs. Fisheries Research, 1995, 22, 155–160.

Examples

```
### Multiplicative model from Aitchison and Silvey (1960)

A = matrix(c(1, 0, 0, 1, 0, 1, 1,
             0, 1, 0, 1, 1, 0, 1,
             0, 0, 1, 0, 1, 1, 1), byrow=TRUE, nrow=3) ## the model matrix

y = c(46,24,7,15,3,4,1) ## the observed data

g.ipf(A, y, 1e-6, "probabilities", "bisection")
g.ipf(A, y, 1e-4, "probabilities", "grid")



### Bait study in swimming crabs, see Kawamura at al. (1995) :

A = matrix(c(1,1,0,1,0,1), 2,3,byrow=TRUE) ## the model matrix

y1 = c(36,2,11)  ## the observed data for Charybdis japonica
y2 = c(71,3,44)  ## the observed data for Portunuspelagicus

## If the model is for intensities, the adjustemnt factor is equal to one;
## the adjustment parameter is redundant.

g.ipf(A, y1, 1e-6, "intensities", "none")
g.ipf(A, y2, 1e-6, "intensities", "none")
```

`grid.update` *Search for the adjustment factor corresponding to the MLE using a grid*

Description

`grid.update` computes the segment that is known to contain the adjustment factor corresponding to the MLE and, using a grid on this segment, finds this adjsutment factor. It is needed only for relational models for probabilities.

Usage

```
grid.update(ModelMx, ObsTbl, tolerance)
```

Arguments

`ModelMx`	an `I` by `J` model matrix of a relational model. Here `I` is the number of observations and `J` is the number of generating subsets.
`ObsTbl`	a vector of observed cell frequencies of length `I`.
`tolerance`	tolerance used in stopping criteria.

Value

`gamma.tilde` the adjustment factor under the precision given by `tolerance`.

`model.tilde` the value returned by `ipf.gamma()` with the adjustment factor `gamma` equal to `gamma.tilde`.

Author(s)

Anna Klimova, Tamas Rudas

References

A. Klimova, Coordinate-Free Exponential Families on Contingency Tables. PhD thesis. Advisers: Tamas Rudas and Thomas Richardson.

Examples

```
### Multiplicative model from Aitchison and Silvey (1960)

A = matrix(c(1, 0, 0, 1, 0, 1, 1,
             0, 1, 0, 1, 1, 0, 1,
             0, 0, 1, 0, 1, 1, 1), byrow=TRUE, nrow=3) ## the model matrix

y = c(46,24,7,15,3,4,1) ## the observed data

grid.update(A, y, 1e-4)


## The model of independence for a 2 by 2 contingency table

A = matrix(c( 1,1,0,0,
              0,0,1,1,
              1,0,1,0,
              0,1,0,1), byrow=TRUE, nrow=4) ## the model matrix

y = c(1,2,3,4)  ## the observed data


grid.update(A, y, 1e-5)
```

`ipf.gamma` *Iterative Proportional Fitting in Relational Models, with a Given Adjustment Factor*

Description

For a given model matrix and a given vector of observed cell frequencies, `ipf.gamma` computes the vector of frequencies whose subset sums are equal to the observed subset sums times the adjustment factor and whose relative frequencies satisfy the multiplicative structure prescribed by the model.

Usage

```
ipf.gamma(ModelMatrix, ObsTable, gamma, tol, estimand)
```

Arguments

- `ModelMatrix` an I by J model matrix of a relational model. I is the number of observations; J is the number of generating subsets.
- `ObsTable` a vector of observed cell frequencies.
- `gamma` an adjustment factor.
- `tol` tolerance used in stopping criteria.
- `estimand` set to "probabilities" in the case of multinomial sampling; set to "intensities" in the case of Poisson sampling.

Value

- `model.matrix` the model matrix.
- `observed.data` the vector of observed cell frequencies.
- `fitted.values` the estimated cell frequencies.
- `model.parameters` the estimated model parameters on the multiplicative scale.

Author(s)

Anna Klimova, Tamas Rudas

References

A.Klimova, T.Rudas, A.Dobra, Relational models for contingency tables. J. Multivariate Anal., 2012, 104, 159–173.

A.Klimova, T.Rudas, Iterative proportional scaling for curved exponential families. Scand. J. Statist., 2015, 42, 832–847.

A. Klimova, Coordinate-Free Exponential Families on Contingency Tables. PhD thesis. Advisers: Tamas Rudas and Thomas Richardson.

Examples

```
## The model of independence for a 2 by 2 contingency table

A = matrix(c( 1,1,0,0,
              0,0,1,1,
              1,0,1,0,
              0,1,0,1), byrow=TRUE, nrow=4) ## the model matrix

y = c(1,2,3,4)  ## the observed data

ipf.gamma(A, y, 1, 1e-6, "intensities")   ## Adjustment factor is set to 1
```

```
ipf.gamma(A, y, 0.5, 1e-6, "intensities")   ## Adjustment factor is set to 0.5
```

---

single.cells *Single Cells under a Relational Model*

Description

The function finds all single cells under a relational model. Such cells appear as the only positive entries in their row and column in the model matrix.

Usage

```
single.cells(ModelMatrix)
```

Arguments

ModelMatrix	a model matrix of a relational model.

Value

the row and column indices of the single cells.

Author(s)

Anna Klimova

Examples

```
G = matrix(c( 2,0,0,0,0,
              0,1,1,0,0,
              0,0,1,1,1), byrow=TRUE, nrow=3) ## a given matrix

single.cells(G)
```

suff.stat — *Sufficient Statistics under a Relational Model*

Description

For an `I` by `J` model matrix of a relational model and a vector of frequencies of the length `I`, the function computes sufficient statistics under the model (subset sums).

Usage

```
suff.stat(ModelMatrix, Table)
```

Arguments

`ModelMatrix`	`ModelMatrix` a model matrix of a relational model.
`Table`	`Table` a vector of frequencies.

Value

a vector of subset sums.

Author(s)

Anna Klimova

References

A.Klimova, T.Rudas, A.Dobra, Relational models for contingency tables. J. Multivariate Anal., 104, 159–173.

Examples

```
### Multiplicative model from Aitchison and Silvey (1960)

A = matrix(c(1, 0, 0, 1, 0, 1, 1,
             0, 1, 0, 1, 1, 0, 1,
             0, 0, 1, 0, 1, 1, 1), byrow=TRUE, nrow=3) ## the model matrix

y = c(46,24,7,15,3,4,1) ## the observed data

suff.stat(A,y)

### pneumonia infection in calves:

A=matrix(c(2,1,0,0,1,1), 2,3,byrow=TRUE) ## the model matrix

y=c(30,63,63)  ## the observed data

suff.stat(A,y)
```

Index