Package 'eMLEloglin'

October 13, 2022

Title Fitting log-Linear Models in Sparse Contingency Tables

Version 1.0.1

Date 2016-11-23

Author Matthew Friedlander

Maintainer Matthew Friedlander <mattyf5@hotmail.com>

Depends lpSolveAPI

Description Log-linear modeling is a popular method for the analysis of contingency table data. When the table is sparse, the data can fall on the boundary of the convex support, and we say that ``the MLE does not exist" in the sense that some parameters cannot be estimated. However, an extended MLE always exists, and a subset of the original parameters will be estimable. The 'eMLEloglin' package determines which sampling zeros contribute to the non-existence of the MLE. These problematic zero cells can be removed from the contingency table and the model can then be fit (as far as is possible) using the glm() function.

License GPL (>= 2)

NeedsCompilation no

Repository CRAN

Date/Publication 2016-11-30 18:31:21

R topics documented:

eMLEloglin *Fitting log-linear models in sparse contingency tables.*

Description

Log-linear modeling is a popular method for the analysis of contingency table data. When the table is sparse, the data can fall on the boundary of the convex support, and we say that "the MLE does not exist" in the sense that some parameters cannot be estimated. However, an extended MLE always exists, and a subset of the original parameters will be estimable. The 'eMLEloglin' package determines which sampling zeros contribute to the non-existence of the MLE. These problematic zero cells can be removed from the contingency table and the model can then be fit using the glm() function.

Details

Package:	eMLEloglin
Type:	Package
Version:	1.0.1
Date:	2016-11-23
License:	GPL-2

The function facial_set identifies sampling zeros that contribute to the non-existence of the MLE.

Author(s)

Author: Matthew Friedlander Maintainer: Matthew Friedlander <mattyf5@hotmail.com>

References

Feinberg, S. E. and Rinaldo, A. (2012). Maximum likelihood estimation in log-linear models. Annals of Statistics, 40: 996-1023.

Friedlander, M. (2016). Fitting log-linear models in sparse contingency tables using the eMLEloglin R package. Preprint. arXiv:1611.07505

Examples

```
data(rochdale)
f <- facial_set (data = rochdale,
       formula = freq ~ a*d + a*e + b*e + c*e + e*f + a*c*g + d*g + f*g + b*d*h)
N <- sum(rochdale$freq)
BIC <- f$maxloglik - f$model.dimension * log(N) / 2
cBIC <- f$maxloglik - f$face.dimension * log(N) / 2
```

`facial_set` *Finds the facial set*

Description

Identifies sampling zeros that contribute to the non-existence of the MLE. These problematic zero cells can be removed from the contingency table and the model can then be fit (as far as is possible) using the glm package.

Usage

```
facial_set (data, formula)
```

Arguments

`data`	A data frame containing the contingency table. All cells must be included in data and the last column must be the cell counts. The number of variables in the contingency table must be between 2 and 8.
`formula`	A model formula.

Value

A list with 7 elements:

`formula`	The model formula used in the function call.
`model.dimension`	The model dimension under Poisson sampling.
`status`	Description of how the algorithm terminated. For debugging purposes.
`iterations`	Number of iterations the algorithm used.
.	
`face`	The contingency table with an appended column denoting whether each cell is part of the facial set (denoted 0) or not (denoted 1). Those cells that are not part of the facial set are cells with an estimated cell mean of 0. These cells cause the non-existence of the MLE and should be removed from the table before the model is fit by the glm function. See the vignette for more details.
`face.dimension`	The dimension of the facial set. Equivalent to the number of log-linear parameters that can be estimated.
`maxloglik`	The maximum of the log likelihood function up to a constant.

Author(s)

Matthew Friedlander

References

Feinberg, S. E. and Rinaldo, A. (2012). Maximum likelihood estimation in log-linear models. Annals of Statistics. 40: 996-1023

Friedlander, M. (2016). Fitting log-linear models in sparse contingency tables using the eMLEloglin R package. Preprint. arXiv:1611.07505

Examples

```
data(rochdale)
f <- facial_set (data = rochdale,
       formula = freq ~ a*d + a*e + b*e + c*e + e*f + a*c*g + d*g + f*g + b*d*h)
N <- sum(rochdale$freq)
BIC <- f$maxloglik - f$model.dimension * log(N) / 2
cBIC <- f$maxloglik - f$face.dimension * log(N) / 2
```

---

rochdale *The rochdale data*

Description

An 8 way sparse contingency table representing the cross classification of N=665 individuals. All 8 classification criteria are binary. The study was conducted to elicit information about factors affecting the pattern of economic life in Rochdale, England. The table is sparse have 165 counts of zero, 217 counts with at most three observations, but also a few large counts with 30 or more observations. The variables are: a. wife economically active (no, yes); b. age of wife >38 (no, yes); c. husband unemployed (no, yes); d. child <=4 (no, yes); e. wife's education, highschool+ (no, yes); f. husband's education, highschool+ (no, yes); g. Asian origin (no, yes); h. other household member working (no, yes).

Usage

```
data(rochdale)
```

Source

Whittaker (1990)

References

[1] Whittaker, J. (1990). Graphical models in applied multivariate statistics. John Wiley & Sons.

Index