Package ‘bayesloglin’

December 27, 2016

Title Bayesian Analysis of Contingency Table Data

Version 1.0.1

Date 2016-12-18

Author Matthew Friedlander

Maintainer Matthew Friedlander <friedla@yorku.ca>

Depends igraph

Description The function MC3() searches for log-linear models with the highest posterior probability. The function gibbsSampler() is a blocked Gibbs sampler for sampling from the posterior distribution of the log-linear parameters. The functions findPostMean() and findPostCov() compute the posterior mean and covariance matrix for decomposable models which, for these models, is available in closed form.

NeedsCompilation yes

Repository CRAN

Date/Publication 2016-12-27 15:28:19

R topics documented:

`bayesloglin` *Bayesian analysis of contingency table data*

Description

Functions for Bayesian model selection and inference for log-linear models.

Details

Package:	bayesloglin
Type:	Package
Version:	1.0
Date:	2016-12-23
License:	GPL-2

The function MC3 searches for log-linear models with the highest posterior probability. The function gibbsSampler is a blocked Gibbs sampler for sampling fronm the posterior distribution of the log-linear parameters. The functions findPostMean and findPostCov compute the posterior mean and covariance matrix for decomposable models which, for these models, is available in closed form.

Author(s)

Author: Matthew Friedlander Maintainer: Matthew Friedlander <friedla@yorku.ca>

References

see vignette

Examples

```
data(czech)
s1 <- MC3 (init = NULL, alpha = 1, iterations = 5,
           replicates = 1, data = czech, mode = "Decomposable")
s2 <- MC3 (init = NULL, alpha = 1, iterations = 5,
            replicates = 1, data = czech, mode = "Graphical")
s3 <- MC3 (init = NULL, alpha = 1, iterations = 5,
            replicates = 1, data = czech, mode = "Hierarchical")
```

`czech` *The Czech autoworkers data*

Description

A 6 way contingency table representing the cross classification of 1841 men. All 6 classification criteria are binary. The variables are (a) smoking, (b) strenuous mental work, (c) strenuous physical work, (d) systolic blood pressure, (e) ratio of beta and alpha lipoproteins and (f) family anamnesis of coronary heart disease.

Usage

```
data(czech)
```

Source

Edwards and Havranek (1985)

References

[1] Edwards, D. E. and Havranek, T. (1985). A fast procedure for model search in multidimensional contingency tables. Biometrika, 72 339-351.

`findPostCov` *Posterior covariance matrix for a decomposable model.*

Description

Computes the posterior covariance matrix of the log-linear parameters, which for decomposable models, is known in closed form.

Usage

```
findPostCov(formula, alpha, data)
```

Arguments

`formula`	A decomposable model formula.
`alpha`	The value of the hyperparameter alpha.
`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 at least 2.

Value

`theta`	An array giving the posterior covariance matrix of the log-linear parameters.

Author(s)

Matthew Friedlander

References

see vignette

Examples

```
data(czech)
formula <- freq ~ b*c + a*c*e + d*e + f
s <- findPostCov (formula, alpha = 1, data = czech)
print(s)
```

`findPostMean` *Posterior covariance matrix for a decomposable model.*

Description

Computes the posterior mean, which for decomposable models, is known in closed form.

Usage

```
findPostMean(formula, alpha, data)
```

Arguments

`formula`	A decomposable model formula.
`alpha`	The value of the hyperparameter alpha.
`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 at least 2.

Value

`theta`	An array giving the posterior mean of the log-linear parameters.

Author(s)

Matthew Friedlander

References

see vignette

Examples

```
data(czech)
formula <- freq ~ b*c + a*c*e + d*e + f
s <- findPostMean (formula, alpha = 1, data = czech)
print(s)
```

gibbsSampler *A blocked Gibbs sampler.*

Description

Generates samples from the posterior distribution of the log-linear parameters.

Usage

```
gibbsSampler (formula, alpha = 1, data, nSamples = 10000, verbose = T)
```

Arguments

`formula`	A model formula.
`alpha`	The value of the hyperparameter alpha.
`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 at least 2.
`nSamples`	Number of iterations of the Gibbs sampler.
`verbose`	Displays current iteration number of the sampler.

Value

`theta`	An array where each row represents a sample from the Posterior distribution of the log-linear parameters. The first 5000 or so samples should be discarded as a burn-in period.

Author(s)

Matthew Friedlander

References

see vignette

Examples

```
data(czech)
formula <- freq ~ a*c + b*c + a*d + a*e + c*e + d*e + f
#s <- gibbsSampler (formula, alpha = 1, data = czech,
#                   nSamples =  15000, verbose = TRUE)
#postMean <- colSums(s[5000:15000,]) / 10000
#postCov <- cov(s[5000:15000,])
#postVar <- diag(postCov)
```

MC3 *Implementation of the MC3 algorithm*

Description

The MC3 algorithm is used to find the log-linear models with the highest posterior probability. The function is capable of searching the space of decomposable, graphical, and hierarchical models.

Usage

```
MC3 (init = NULL, alpha = 1, iterations = 5000, replicates = 1, data,
            mode = c("Hierarchical", "Graphical", "Decomposable"))
```

Arguments

`init`	A starting list of models for the MOSS algorithm. If not NULL, this list should consist of R formulas.
`alpha`	The value of the hyperparameter alpha.
`iterations`	Number of iterations of the MC3 algorithm
`replicates`	The number of instances the MC3 algorithm will be run. The top models are culled from the results of all the replicates.
`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 at least 3.
`mode`	The scope of the loglinear model search. The search can be over the space of decomposable, graphical, or hierarchical models.

Value

A data frame with the top models found and their log posterior probability up to a constant.

Author(s)

Matthew Friedlander

References

see vignette

Examples

```
data(czech)
s1 <- MC3 (init = NULL, alpha = 1, iterations = 5,
           replicates = 1, data = czech,
           mode = "Decomposable")
s2 <- MC3 (init = NULL, alpha = 1, iterations = 5,
           replicates = 1, data = czech,
```

```
            mode = "Graphical")
s3 <- MC3 (init = NULL, alpha = 1, iterations = 5,
            replicates = 1, data = czech,
            mode = "Hierarchical")
```

Index