Package ‘mlogitBMA’

February 20, 2015

Type Package
Title Bayesian Model Averaging for Multinomial Logit Models
Version 0.1-6
Date 2013-12-11
Depends R (>= 2.9.0), BMA, abind, maxLik
Suggests mlogit
Author Hana Sevcikova, Adrian Raftery
Maintainer Hana Sevcikova <hanas@uw.edu>
Description Provides a modified function bic.glm of the BMA package that can be applied to multinomial logit (MNL) data. The data is converted to binary logit using the Begg & Gray approximation. The package also contains functions for maximum likelihood estimation of MNL.
License GPL (>= 2)
NeedsCompilation yes
Repository CRAN
Date/Publication 2013-12-11 14:55:55

R topics documented:

mlogitBMA-package .................................................. 2
bic.mlogit .......................................................... 3
estimate.mlogit ...................................................... 5
heating ............................................................... 7
mlogit2logit .......................................................... 9
mnl.spec ............................................................ 10
summary.bic.mlogit .................................................. 12
summary.mnl ........................................................... 13
summary.mnl.spec ..................................................... 13

Index 15
mlogitBMA-package

Bayesian Model Averaging for Multinomial Logit Models

Description

Provides a modified function bic.glm of the BMA package that can be applied to multinomial logit (MNL) data. The data is converted to binary logit using the Begg & Gray approximation. The package also contains functions for maximum likelihood estimation of MNL models.

Details

Package: mlogitBMA
Version: 0.1-6
Date: 2012-12-11
Depends: R (>= 2.9.0), BMA, abind, maxLik

The main function of the package is bic.mlogit which runs the Bayesian Model Averaging on multinomial logit data. Results can be explored using summary.bic.mlogit, imageplot.mlogit, or plot.bic.mlogit functions.

An MNL estimation of a single model can be done using estimate.mlogit. Use summary.mnl to view its results.

Author(s)

Hana Sevcikova, Adrian Raftery
Maintainer: Hana Sevcikova <hanas@uw.edu>

References


See Also

bic.glm
**bic.mlogit**

**Bayesian Model Averaging for Multinomial Logit Models**

**Description**

Using the methodology of Bayesian Model Averaging in the BMA package, the variable selection problem is applied to multinomial logit models in which coefficients can be estimated relative to a base alternative.

**Usage**

```r
bic.mlogit(f, data, choices = NULL, base.choice = 1,
      varying = NULL, sep = ".", approx=TRUE,
      include.intercepts = TRUE, verbose = FALSE, ...)
```

**Arguments**

- `f`  
  Formula as described in Details of `mnl.spec`.

- `data`  
  Data frame containing the variables of the model. There should be one record for each individual. Alternative-specific variables occupy single column per alternative.

- `choices`  
  Vector of names of alternatives. If it is not given, it is determined from the response column of the data frame. Values of this vector should match or be a subset of those in the response column. If it is a subset, data is reduced to contain only observations whose choice is contained in `choices`.

- `base.choice`  
  Index of the base alternative within the vector `choices`.

- `varying`  
  Indices of variables within `data` that are alternative-specific.

- `sep`  
  Separator of variable name and alternative name in the 'varying' variables.

- `approx`  
  Logical. If TRUE, the function uses approximate likelihoods as they come out of the Begg & Gray approximation. If FALSE, the MNL maximum likelihood estimation is used in the last step of the model selection procedure. Note that this can significantly increase the run-time, see Details below.

- `include.intercepts`  
  Logical controlling if alternative specific constants should always be included in the selected models. It only has an effect if the formula `f` contains the intercept, i.e. it does not contain `-1`. See Details below.

- `verbose`  
  Logical switching log messages on and off.

- `...`  
  Additional arguments passed to the `bic.glm` function of the BMA package.

**Details**

The function converts the given multinomial data into a combination of binary logistic data, as proposed in Yeung et al. (2005). It requires that the model can be specified as a set of equations of which one is considered as the base equation. If variables are included that vary over alternatives,
they are normalized by subtracting the values corresponding to the base alternative. Details of the conversion algorithm are described in the vignette of this package, see vignette(’conversion’).

The function then applies the \texttt{bic.glm} function of the \texttt{BMA} package on the converted data by using the Begg & Gray (1984) approximation. In the last step of the variable selection procedure, if \texttt{approx} is \texttt{FALSE}, the maximum likelihood estimation (MLE) is applied to all selected models and the Bayesian Information Criterium (BIC) is recomputed using the log-likelihood of the full multinomial logistic regression model. Note that this step can be computationally very expensive. We suggest when using this option, set the \texttt{verbose} argument to \texttt{TRUE} to follow the computation progress. Note that one can use the \texttt{estimate.mlogit} function on the resulting object which performs the MLE on selected models only.

The \texttt{BMA} functions always include the intercept which in the MNL settings corresponds to the alternative specific constant (asc) of the second alternative (relative to the base alternative). If \texttt{include.intercepts=TRUE} (default), asc for all the remaining alternatives are also always included in the selected models. If it is set to \texttt{FALSE}, the asc of the remaining alternatives (i.e. third and higher) are treated as ordinary variables, i.e candidates for selection as well as exclusion.

\textbf{Value}

The function returns an object of class \texttt{bic.mlogit} containing the following components:

\begin{description}
\item[bic.glm] Object of class \texttt{bic.glm} which results from applying BMA on the binary logistic data.
\item[bin.logit] List with results from the \texttt{mlogit2logit} function.
\item[spec] Object of class \texttt{mnl.spec} containing the MNL specification of the full model.
\item[bma.specifications] List of objects of class \texttt{mnl.spec} containing specifications for each selected model.
\item[approx] Value of the \texttt{approx} argument.
\end{description}

\textbf{Author(s)}

Hana Sevcikova, Adrian Raftery

\textbf{References}


\textbf{See Also}

\texttt{bic.glm}, \texttt{summary.bic.mlogit}, \texttt{imageplot.mlogit}, \texttt{estimate.mlogit}.
Examples

data('heating')
res <- bic.mlogit(depvar ~ ic + oc + income + rooms, heating, choices=1:5,
                  varying=3:12, verbose=TRUE, approx=FALSE, sep = ' ')
summary(res)
imageplot.mlogit(res)
plot(res)

# use approximate BMA and estimate the models afterwards
res <- bic.mlogit(depvar ~ ic + oc | income + rooms, heating, choices=1:5,
                  varying=3:12, verbose=TRUE, approx=TRUE, sep = '')
summary(res)
estimate.mlogit(res, heating)

estimate.mlogit Multinomial Logit Estimation

Description

Maximum likelihood estimation of coefficients of one or more multinomial logit models.

Usage

## S3 method for class 'formula'
estimate.mlogit(f, data, method = "BHHH",
                  choices = NULL, base.choice = 1,
                  varying = NULL, sep = ".", ...)  

## S3 method for class 'mnl.spec'
estimate.mlogit(object, data, method='BHHH', ...)  

## S3 method for class 'bic.mlogit'
estimate.mlogit(object, ...)  

## S3 method for class 'list'
estimate.mlogit(object, data, verbose=TRUE, ...)

Arguments

f Formula as described in Details of mnl.spec.

object An object of class mnl.spec containing the model specification, or an object of
class bic.mlogit, or a list of objects of class mnl.spec.

data Data frame containing the variables of the model.

method Estimation method passed to the maxLik function of the maxLik package. Avail-
able methods are "Newton-Raphson", "BFGS", "BHHH", "SANN" or "NM".
choices Vector of names of alternatives. If it is not given, it is determined from the response column of the data frame. Values of this vector should match or be a subset of those in the response column. If it is a subset, data is reduced to contain only observations whose choice is contained in choices.

base.choice Index of the base alternative within the vector choices.

varying Indices of variables within data that are alternative-specific.

sep Separator of variable name and alternative name in the ‘varying’ variables.

verbose Logical switching log messages on and off.

... Arguments passed to the underlying optimization routine in optim. Note that arguments data and method can be also passed to estimate.mlogit.bic.mlogit and estimate.mlogit.list.

Details

The data are expected to be in the ‘wide’ format (using the terminology of the reshape function). There should be one record for each individual. Alternative-specific variables occupy single column per alternative. The given optimization routine is called for the multinomial data, starting from the coefficients being all zeros.

Function estimate.mlogit.bic.mlogit invokes as many estimations as there are models selected in the bic.mlogit object. Function estimate.mlogit.list invokes an estimation for each specification included in the object argument.

Value

Functions estimate.mlogit.formula and estimate.mlogit.mnl.spec return an object of class mnl. Functions estimate.mlogit.bic.mlogit and estimate.mlogit.list return a list of such objects with each element corresponding to one specification. An object of class mnl contains the following components:

coefficients The estimated coefficients.
logLik Maximum log-likelihood.
logLik0 Null log-likelihood.
aic Akaike Information Criterion.
bic Bayesian Information Criterion.
iter Number of iterations.
hessian The Hessian at the maximum.
gradient The last gradient value.
fitted.values The MNL probabilities computed with the estimated parameters.
residuals Difference between observed values and fitted values.
specification The corresponding mnl.spec object.
convergence Convergence statistics.
method Estimation method.
time Time needed for the estimation.
code Code returned by the maxLik function.
message Message describing the code.
last.step List describing the last unsuccessful step if code=3 (see maxLik).
heating

Author(s)
Hana Sevcikova

References

See Also
summary.mnl, mnl.spec, reshape, maxLik

Examples

```r
data(heating)
est <- estimate.mlogit(depvar ~ ic + oc, heating, choices=1:5,
                        varying=c(3:12, 20:24), sep='')
summary(est)
```

---

heating  

**Heating Dataset**

Description
Kenneth Trains dataset containing data on choice of heating system in California houses.

Usage

```r
data(heating)
```

Format
A data frame with 900 observations on the following 19 variables.

- idcase  Observation number.
- depvar  Identifies the chosen alternative (1-5).
- ic1  Installation cost for a gas central system.
- ic2  Installation cost for a gas room system.
- ic3  Installation cost for a electric central system.
- ic4  Installation cost for a electric room system.
- ic5  Installation cost for a heat pump.
- oc1  Annual operating cost for a gas central system.
- oc2  Annual operating cost for a gas room system.
- oc3  Annual operating cost for a electric central system.
- oc4  Annual operating cost for a electric room system.
- oc5  Annual operating cost for a heat pump.
income  Annual income of the household.
agehed  Age of the household head.
rooms   Number of rooms in the house.
ncost1  Identifies whether the house is in the northern coastal region.
scost1  Identifies whether the house is in the southern coastal region.
mountain Identifies whether the house is in the mountain region.
valley  Identifies whether the house is in the central valley region.

Details

The observations consist of single-family houses in California that were newly built and had central air-conditioning. The choice is among heating systems. Five types of systems are considered to have been possible:

(1) gas central, (2) gas room, (3) electric central, (4) electric room, (5) heat pump.

For these data, the costs were calculated as the amount the system would cost if it were installed in the house, given the characteristics of the house (such as size), the price of gas and electricity in the house location, and the weather conditions in the area (which determine the necessary capacity of the system and the amount it will be run.) These cost are conditional on the house having central air-conditioning. (That is why the installation cost of gas central is lower than that for gas room: the central system can use the air-conditioning products that have been installed.)

Note

This help file was created using Kenneth Trains description of the dataset, see Source.

Source

http://elsa.berkeley.edu/~train/distant.html

References


Examples

data(heating)
head(heating)
Converting Multinomial Logit Data into Binary Logit Data

Description

Converts multinomial logit data into a combination of several binary logit data sets, in order to analyze it via the Begg & Gray approximation using a binary logistic regression.

Usage

mlogit2logit(f, data, choices = NULL, base.choice = 1, varying = NULL, sep = ".")

Arguments

f Formula as described in Details of mnl.spec.
data Data frame containing the variables of the model.
choices Vector of names of alternatives. If it is not given, it is determined from the response column of the data frame. Values of this vector should match or be a subset of those in the response column. If it is a subset, data is reduced to contain only observations whose choice is contained in choices.
base.choice Index of the base alternative within the vector choices.
varying Indices of variables within data that are alternative-specific.
sep Separator of variable name and alternative name in the ‘varying’ variables.

Details

Details of the conversion algorithm are described in the vignette of this package, see vignette('conversion').

Value

List with components:
data Converted data set.
formula Formula to be used with the converted data set.
nobs Number of observations in the original data set.
z.index Index of all Z columns within data (see vignette for details), i.e. columns that correspond to alternative specific constants.
z.names Names of the Z columns.
zcols List in which each element corresponds to any of the data columns that involve Z, which is either Z itself or an interaction between a variable and Z, (see vignette). The value of such element is a vector with the components 'name': either Z itself, or name of the corresponding X or U variable with which Z interacts; 'choice': which alternative it belongs to; 'intercept': logical determining if it is an alternative specific constant.
choices Vector of names of the alternatives.

choice.main.intercept Index of alternative within choices that corresponds to the main intercept of the binary logistic model.

Note

This function is called from within the \texttt{bic.mlogit} and thus usually will not need to be called explicitly.

Author(s)

Hana Sevcikova

References


See Also

\texttt{mnl.spec}

Examples

data(heating)
bin.data <- mlogit2logit(depvar ~ ic + oc, heating, choices=1:5,
                       varying=3:12, sep="")
bin.glm <- glm(bin.data$formula, 'binomial', data=bin.data$data)
summary(bin.glm)

mnl.spec \hspace{1cm} \textit{Specification Object of a Multinomial Logit Model}

Description

Using a formula and data, create a specification object of a multinomial logit model.

Usage

\texttt{mnl.spec(f, data, choices = NULL, base.choice = 1,}
\texttt{varying = NULL, sep = ".")}
Arguments

\textbf{f} \quad \text{Formula (see Details below).}

\textbf{data} \quad \text{Data frame containing the variables in the model. It should be in the ‘wide’ format (using the terminology of the \texttt{reshape} function), i.e. there is one record for each individual and alternative-specific variables occupy single column per alternative.}

\textbf{choices} \quad \text{Vector of names of alternatives. If it is not given, it is determined from the response column of the data frame. Values of this vector should match or be a subset of those in the response column.}

\textbf{base.choice} \quad \text{Index of the base alternative within the vector \texttt{choices}.}

\textbf{varying} \quad \text{Indices of variables within \texttt{data} that are alternative-specific.}

\textbf{sep} \quad \text{Separator of variable name and alternative name in the ‘varying’ variables.}

Details

The formula \textit{f} is of the form \texttt{response} \sim x_1 + x_2 \mid y_1 + y_2. Coefficients for variables in the first part of the formula (i.e. before \texttt{'}|\texttt{'}), here \texttt{x1} and \texttt{x2}, are forced to be the same for all alternatives. Variables in the second part of the formula (i.e. after \texttt{'}|\texttt{'}), here \texttt{y1} and \texttt{y2}, have different coefficients for different alternatives. Either part of the formula can be omitted. Alternative specific constants (asc) are included automatically. To exclude asc, use -1 in the first part. The equation of the base alternative is always set to 0.

Value

An object of class \texttt{mnl.spec} containing the following elements:

\textbf{response} \quad \text{Name of the response variable.}
\textbf{choices} \quad \text{Vector of alternatives.}
\textbf{base.choice} \quad \text{Index of the base alternative within \texttt{choices}.}
\textbf{variable.used} \quad \text{Matrix of size number of choices x number of variables. Each value is logical determining if the variable is used in that choice equation.}
\textbf{same.coefs} \quad \text{Logical vector of size number of variables. It determines if that variable has the same coefficient for all alternatives.}
\textbf{full.var.names} \quad \text{Matrix of the same shape as \texttt{variable.used}. It contains names of variables in its alternative-specific form.}
\textbf{varying.names} \quad \text{Vector of variable names specified by the \texttt{varying} vector that are used in the specification.}
\textbf{intercepts} \quad \text{Logical vector of size number of choices determining in which equation asc is used.}
\textbf{sep} \quad \text{Separator of variable name and alternative name in the ‘varying’ variables.}
\textbf{frequency} \quad \text{Table of frequencies for each choice in the \texttt{choices} vector computed from the data.}

Author(s)

Hana Sevcikova
### Description

Summarizes and plots results of the `bic.mlogit` function.

### Usage

```r
## S3 method for class 'bic.mlogit'
summary(object, ...)  
## S3 method for class 'bic.mlogit'
plot(x, ...)  
imageplot.mlogit(x, ...)
```

### Arguments

- `object, x`: Object of class `bic.mlogit`.
- `...`: Arguments passed to the underlying functions.

### Details

- `summary` prints a summary of `object`, using the **BMA** function `summary.bic.glm`. It also prints a summary of the model specification, using `summary.mnl.spec`.
- `plot` produces a plot of the posterior distribution of the coefficients produced by model averaging. It uses the **BMA** function `plot.bic.glm`.
- `imageplot.mlogit` creates an image of the selected models, using the **BMA** function `imageplot.bma`.

### Author(s)

Hana Sevcikova
See Also

bicNmlogit

Examples

# See example in bic.mlogit
Author(s)
Hana Sevcikova

See Also
mnl.spec

Examples

data(heating)
spec <- mnl.spec(depvar ~ ic | oc, heating, varying=3:12, sep='')
summary(spec)
Index

+ Topic **classes**
  - bic.mlogit, 3
  - estimate.mlogit, 5
  - mnl.spec, 10
+ Topic **datasets**
  - heating, 7
+ Topic **hplot**
  - summary.bic.mlogit, 12
+ Topic **htest**
  - estimate.mlogit, 5
+ Topic **manip**
  - mlogit2logit, 9
+ Topic **models**
  - bic.mlogit, 3
  - mnl.spec, 10
+ Topic **optimize**
  - estimate.mlogit, 5
+ Topic **package**
  - mlogitBMA-package, 2
+ Topic **print**
  - summary.bic.mlogit, 12
  - summary.mnl, 13
  - summary.mnl.spec, 13
+ Topic **regression**
  - bic.mlogit, 3

bic.glm, 2–4
bic.mlogit, 2, 3, 5, 6, 10, 12, 13

estimate.mlogit, 2, 4, 5

heating, 7

imageplot.bma, 12
imageplot.mlogit, 2, 4
imageplot.mlogit (summary.bic.mlogit), 12

maxLik, 6, 7
mlogit2logit, 4, 9
mlogitBMA (mlogitBMA-package), 2

mlogitBMA-package, 2
mnl, 13
mnl (estimate.mlogit), 5
mnl.spec, 3–7, 9, 10, 10, 13, 14
plot.bic.glm, 12
plot.bic.mlogit, 2
plot.bic.mlogit (summary.bic.mlogit), 12
reshape, 6, 7, 11
summary.bic.glm, 12
summary.bic.mlogit, 2, 4, 12
summary.mnl, 2, 7, 13
summary.mnl.spec, 12, 13

15