Package 'mlogitBMA'

July 23, 2025

Type Package

Title Bayesian Model Averaging for Multinomial Logit Models

Version 0.1-9

Date 2024-10-16

Depends R (>= 2.9.0), BMA, abind, maxLik

Suggests mlogit

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

Author Hana Sevcikova [aut, cre],
Adrian Raftery [aut]

Maintainer Hana Sevcikova <hanas@uw.edu>

Repository CRAN

Date/Publication 2024-10-17 16:30:02 UTC

Contents

`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

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

Begg, C.B., Gray, R. (1984) Calculation of polychotomous logistic regression parameters using individualized regressions. Biometrika **71**, 11–18.

Raftery, A.E. (1995) Bayesian model selection in social research (with Discussion). Sociological Methodology 1995 (Peter V. Marsden, ed.), 111–196, Cambridge, Mass.: Blackwells.

Train, K.E. (2003) Discrete Choice Methods with Simulation. Cambridge University Press.

Yeung, K.Y., Bumgarner, R.E., Raftery, A.E. (2005) Bayesian model averaging: development of an improved multi-class, gene selection and classification tool for microarray data. Bioinformatics **21** (10), 2394–2402.

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

```
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 `bic.glm` function of the **BMA** package on the converted data by using the Begg & Gray (1984) approximation. In the last step of the variable selection procedure, if `approx` is `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 `verbose` argument to `TRUE` to follow the computation progress. Note that one can use the `estimate.mlogit` function on the resulting object which performs the MLE on selected models only.

The **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 `include.intercepts=TRUE` (default), asc for all the remaining alternatives are also always included in the selected models. If it is set to `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.

Value

The function returns an object of class `bic.mlogit` containing the following components:

`bic.glm`	Object of class `bic.glm` which results from applying BMA on the binary logistic data.
`bin.logit`	List with results from the `mlogit2logit` function.
`spec`	Object of class `mnl.spec` containing the MNL specification of the full model.
`bma.specifications`	List of objects of class `mnl.spec` containing specifications for each selected model.
`approx`	Value of the `approx` argument.

Author(s)

Hana Sevcikova, Adrian Raftery

References

Begg, C.B., Gray, R. (1984) Calculation of polychotomous logistic regression parameters using individualized regressions. Biometrika **71**, 11–18.

Yeung, K.Y., Bumgarner, R.E., Raftery, A.E. (2005) Bayesian model averaging: development of an improved multi-class, gene selection and classification tool for microarray data. Bioinformatics **21** (10), 2394–2402.

See Also

`bic.glm`, `summary.bic.mlogit`, `imageplot.mlogit`, `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. Available 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 Criterium.
bic	Bayesian Information Criterium.
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).

Author(s)

Hana Sevcikova

References

Train, K.E. (2003) Discrete Choice Methods with Simulation. Cambridge University Press.

See Also

summary.mnl, mnl.spec, reshape, maxLik

Examples

```
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

```
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.

`mountn` 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

Train, K.E. (2003) Discrete Choice Methods with Simulation. Cambridge University Press.

Examples

```
data(heating)
head(heating)
```

mlogit2logit *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 `bic.mlogit` and thus usually will not need to be called explicitly.

Author(s)

Hana Sevcikova

References

Begg, C.B., Gray, R. (1984) Calculation of polychotomous logistic regression parameters using individualized regressions. Biometrika **71**, 11–18.

Yeung, K.Y., Bumgarner, R.E., Raftery, A.E. (2005) Bayesian model averaging: development of an improved multi-class, gene selection and classification tool for microarray data. Bioinformatics **21** (10), 2394–2402.

See Also

`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 *Specification Object of a Multinomial Logit Model*

Description

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

Usage

```
mnl.spec(f, data, choices = NULL, base.choice = 1,
         varying = NULL, sep = ".")
```

Arguments

f	Formula (see Details below).
data	Data frame containing the variables in the model. It should be in the 'wide' format (using the terminology of the reshape function), i.e. there is one record for each individual and 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.
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

The formula f is of the form response ~ x1 + x2 | y1 + y2. Coefficients for variables in the first part of the formula (i.e. before '|'), here x1 and x2, are forced to be the same for all alternatives. Variables in the second part of the formula (i.e. after '|'), here y1 and 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 mnl.spec containing the following elements:

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

Author(s)

Hana Sevcikova

See Also

summary.mnl.spec

Examples

```
data(heating)
spec <- mnl.spec(depvar ~ ic + oc + income, heating, varying=3:12, sep='')
summary(spec)
spec <- mnl.spec(depvar ~ oc-1 | ic, heating, varying=3:12, sep='')
summary(spec)
```

---

summary.bic.mlogit *Summary and Plotting Functions*

Description

Summarizes and plots results of the bic.mlogit function.

Usage

```
## 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

bic.mlogit

Examples

```
# See example in bic.mlogit
```

`summary.mnl` *Summary for Results of a Multinomial Logit Estimation*

Description

Gives a summary for an object of class `mnl` which contains results of a multinomial logit estimation.

Usage

```
## S3 method for class 'mnl'
 summary(object, ...)
```

Arguments

`object`	Object of class `mnl`
`...`	Not used.

Author(s)

Hana Sevcikova

`summary.mnl.spec` *Summary for a Specification Object*

Description

Prints summary for a specification object of a multinomial logit model.

Usage

```
## S3 method for class 'mnl.spec'
 summary(object, ...)
```

Arguments

`object`	Object of class `mnl.spec`.
`...`	Not used.

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