Package ‘MNP’
February 19, 2015

Version 2.6-4
Date 2013-06-09
Title R Package for Fitting the Multinomial Probit Model
Author Kosuke Imai <kimai@princeton.edu>, David A. van Dyk <dvd@uci.edu>.
Maintainer Kosuke Imai <kimai@princeton.edu>
Depends R (>= 2.1), MASS, utils
Description MNP is a publicly available R package that fits the Bayesian multinomial probit model via Markov chain Monte Carlo. The multinomial probit model is often used to analyze the discrete choices made by individuals recorded in survey data. Examples where the multinomial probit model may be useful include the analysis of product choice by consumers in market research and the analysis of candidate or party choice by voters in electoral studies. The MNP software can also fit the model with different choice sets for each individual, and complete or partial individual choice orderings of the available alternatives from the choice set. The estimation is based on the efficient marginal data augmentation algorithm that is developed by Imai and van Dyk (2005). "A Bayesian Analysis of the Multinomial Probit Model Using the Data Augmentation," Journal of Econometrics, Vol. 124, No. 2 (February), pp. 311-334. Detailed examples are given in Imai and van Dyk (2005). "MNP: R Package for Fitting the Multinomial Probit Model." Journal of Statistical Software, Vol. 14, No. 3 (May), pp. 1-32.

LazyLoad yes
LazyData yes
License GPL (>= 2)
URL http://imai.princeton.edu/software/MNP.html
NeedsCompilation yes
Repository CRAN
Date/Publication 2013-06-09 20:16:42
**Description**

`coef.mnp` is a function which extracts multinomial probit model coefficients from objects returned by `mnp`. `coefficients.mnp` is an alias for it. `coef` method for class `mnp`.

**Usage**

```r
## S3 method for class 'mnp'
coef(object, subset = NULL, ...)
```

**Arguments**

- `object`: An output object from `mnp`.
- `subset`: A scalar or a numerical vector specifying the row number(s) of `param` in the output object from `mnp`. If specified, the posterior draws of coefficients for those rows are extracted. The default is `NULL` where all the posterior draws are extracted.
- `...`: Further arguments passed to or from other methods.

**Value**

`coef.mnp` returns a matrix (when a numerical vector or `NULL` is specified for `subset` argument) or a vector (when a scalar is specified for `subset` argument) of multinomial probit model coefficients.

**Author(s)**

Kosuke Imai, Department of Politics, Princeton University <kimai@princeton.edu>

**See Also**

`mnp`, `cov.mnp`; MNP home page at [http://imai.princeton.edu/research/MNP.html](http://imai.princeton.edu/research/MNP.html)
cov.mnp

---

*Extract Multinomial Probit Model Covariance Matrix*

**Description**

cov.mnp is a function which extracts the posterior draws of covariance matrix from objects returned by mnp.

**Usage**

cov.mnp(object, subset = NULL, ...)

**Arguments**

- `object`: An output object from mnp.
- `subset`: A scalar or a numerical vector specifying the row number(s) of `param` in the output object from mnp. If specified, the posterior draws of covariance matrix for those rows are extracted. The default is `NULL` where all the posterior draws are extracted.
- `...`: further arguments passed to or from other methods.

**Value**

When a numerical vector or `NULL` is specified for `subset` argument, cov.mnp returns a three dimensional array where the third dimension indexes posterior draws. When a scalar is specified for `subset` argument, cov.mnp returns a matrix.

**Author(s)**

Kosuke Imai, Department of Politics, Princeton University <kimai@princeton.edu>

**See Also**

mnp, coef.mnp; MNP home page at [http://imai.princeton.edu/research/MNP.html](http://imai.princeton.edu/research/MNP.html)

detergent

---

*Detergent Brand Choice*

**Description**

This dataset gives the laundry detergent brand choice by households and the price of each brand.

**Usage**

data(detergent)
Format

A data frame containing the following 7 variables and 2657 observations.

choice factor a brand chosen by each household
TidePrice numeric log price of Tide
WiskPrice numeric log price of Wisk
EraPlusPrice numeric log price of EraPlus
SurfPrice numeric log price of Surf
SoloPrice numeric log price of Solo
AllPrice numeric log price of All

References


Description

This dataset gives voters’ preferences of political parties in Japan on the 0 (least preferred) - 100 (most preferred) scale. It is based on the 1995 survey data of 418 individual voters. The data also include the sex, education level, and age of the voters. The survey allowed voters to choose among four parties: Liberal Democratic Party (LDP), New Frontier Party (NFP), Sakigake (SKG), and Japanese Communist Party (JCP).

Usage

data(japan)

Format

A data frame containing the following 7 variables for 418 observations.

LDP numeric preference for Liberal Democratic Party 0 - 100
NFP numeric preference for New Frontier Party 0 - 100
SKG numeric preference for Sakigake 0 - 100
JCP numeric preference for Japanese Communist Party 0 - 100
gender factor gender of each voter male or female
education numeric levels of education for each voter
age numeric age of each voter
Description

`mnp` is used to fit (Bayesian) multinomial probit model via Markov chain Monte Carlo. `mnp` can also fit the model with different choice sets for each observation, and complete or partial ordering of all the available alternatives. The computation uses the efficient marginal data augmentation algorithm that is developed by Imai and van Dyk (2005a).

Usage

```r
mnp(formula, data = parent.frame(), choiceX = NULL, cxnames = NULL, base = NULL, latent = FALSE, invcdf = FALSE, trace = TRUE, n.draws = 5000, p.var = "Inf", p.df = n.dim+1, p.scale = 1, coef.start = 0, cov.start = 1, burnin = 0, thin = 0, verbose = FALSE)
```

Arguments

- **formula**: A symbolic description of the model to be fit specifying the response variable and covariates. The formula should not include the choice-specific covariates. Details and specific examples are given below.
- **data**: An optional data frame in which to interpret the variables in formula and choiceX. The default is the environment in which `mnp` is called.
- **choiceX**: An optional list containing a matrix of choice-specific covariates for each category. Details and examples are provided below.
- **cxnames**: A vector of the names for the choice-specific covariates specified in choiceX. The details and examples are provided below.
- **base**: The name of the base category. For the standard multinomial probit model, the default is the lowest level of the response variable. For the multinomial probit model with ordered preferences, the default base category is the last column in the matrix of response variables.
- **latent**: logical. If TRUE, then the latent variable W will be returned. See Imai and van Dyk (2005) for the notation. The default is FALSE.
- **invcdf**: logical. If TRUE, then the inverse cdf method is used for truncated normal sampling. If FALSE, then the rejection sampling method is used. The default is FALSE.
- **trace**: logical. If TRUE, then the trace of the variance covariance matrix is set to a constant (here, it is equal to n.dim) instead of setting its first diagonal element to 1. The former avoids the arbitrariness of fixing one particular diagonal element in order to achieve identification (see Burgette and Nordheim, 2009).
- **n.draws**: A positive integer. The number of MCMC draws. The default is 5000.
p.var  A positive definite matrix. The prior variance of the coefficients. A scalar input can set the prior variance to the diagonal matrix whose diagonal element is equal to that value. The default is "Inf", which represents an improper noninformative prior distribution on the coefficients.

p.df  A positive integer greater than n.dim-1. The prior degrees of freedom parameter for the covariance matrix. The default is n.dim+1, which is equal to the total number of alternatives.

p.scale  A positive definite matrix. When trace = FALSE, its first diagonal element is set to 1 if it is not equal to 1 already. The prior scale matrix for the covariance matrix. A scalar input can be used to set the scale matrix to a diagonal matrix with diagonal elements equal to the scale input value. The default is 1.

coeff.start  A vector. The starting values for the coefficients. A scalar input sets the starting values for all the coefficients equal to that value. The default is 0.

cov.start  A positive definite matrix. When trace = FALSE, its first diagonal element is set to 1 if it is not equal to 1 already. The starting values for the covariance matrix. A scalar input can be used to set the starting value to a diagonal matrix with diagonal elements equal to the scalar input value. The default is 1.

burnin  A positive integer. The burnin interval for the Markov chain; i.e., the number of initial Gibbs draws that should not be stored. The default is 0.

thin  A positive integer. The thinning interval for the Markov chain; i.e., the number of Gibbs draws between the recorded values that are skipped. The default is 0.

verbose  logical. If TRUE, helpful messages along with a progress report of the Gibbs sampling are printed on the screen. The default is FALSE.

Details

To fit the multinomial probit model when only the most preferred choice is observed, use the syntax for the formula, y ~ x1 + x2, where y is a factor variable indicating the most preferred choice and x1 and x2 are individual-specific covariates. The interactions of individual-specific variables with each of the choice indicator variables will be fit.

To specify choice-specific covariates, use the syntax, choiceX=list(A=cbind(z1, z2), B=cbind(z3, z4), C=cbind(z5, where A, B, and C represent the choice names of the response variable, and z1 and z2 are each vectors of length n that record the values of the two choice-specific covariates for each individual for choice A, likewise for z3, ..., z6. The corresponding variable names via cXnames=c("price", "quantity") need to be specified, where price refers to the coefficient name for z1, z3, and z5, and quantity refers to that for z2, z4, and z6.

If the choice set varies from one observation to another, use the syntax, cbind(y1, y2, y3) ~ x1 + x2, in the case of a three choice problem, and indicate unavailable alternatives by NA. If only the most preferred choice is observed, y1, y2, and y3 are indicator variables that take on the value one for individuals who prefer that choice and zero otherwise. The last column of the response matrix, y3 in this particular example syntax, is used as the base category.

To fit the multinomial probit model when the complete or partial ordering of the available alternatives is recorded, use the same syntax as when the choice set varies (i.e., cbind(y1, y2, y3, y4) ~ x1 + x2). For each observation, all the available alternatives in the response variables should be numerically ordered in terms of preferences such as 1 2 3 4. Ties are allowed. The missing values in the response variable should be denoted by NA. The software will impute these missing values using the
specified covariates. The resulting uncertainty estimates of the parameters will properly reflect the amount of missing data. For example, we expect the standard errors to be larger when there is more missing data.

**Value**

An object of class `mnp` containing the following elements:

- **param**
  - A matrix of the Gibbs draws for each parameter; i.e., the coefficients and covariance matrix. For the covariance matrix, the elements on or above the diagonal are returned.

- **call**
  - The matched call.

- **x**
  - The matrix of covariates.

- **y**
  - The vector or matrix of the response variable.

- **w**
  - The three dimensional array of the latent variable, W. The first dimension represents the alternatives, and the second dimension indexes the observations. The third dimension represents the Gibbs draws. Note that the latent variable for the base category is set to 0, and therefore omitted from the output.

- **alt**
  - The names of alternatives.

- **n.alt**
  - The total number of alternatives.

- **base**
  - The base category used for fitting.

- **invcdf**
  - The value of `invcdf`.

- **p.var**
  - The prior variance for the coefficients.

- **p.df**
  - The prior degrees of freedom parameter for the covariance matrix.

- **p.scale**
  - The prior scale matrix for the covariance matrix.

- **burnin**
  - The number of initial burnin draws.

- **thin**
  - The thinning interval.

**Author(s)**

Kosuke Imai, Department of Politics, Princeton University <kimai@princeton.edu>, http://imai.princeton.edu; David A. van Dyk, Department of Statistics, University of California, Irvine <dvd@uci.edu>, http://www.ics.uci.edu/~dvd.

**References**


predict.mnp

See Also
coeff.mnp, cov.mnp, predict.mnp, summary.mnp; MNP home page at http://imai.princeton.edu/research/MNP.html

Examples

###
### NOTE: this example is not fully analyzed. In particular, the
### convergence has not been assessed. A full analysis of these data
### sets appear in Imai and van Dyk (2005b).
###

## load the detergent data
data(detergent)
## run the standard multinomial probit model with intercepts and the price
res1 <- mnp(choice ~ 1, choiceX = list(Surf=SurfPrice, Tide=TidePrice,
Wisk=WiskPrice, EraPlus=EraPlusPrice,
Solo=SoloPrice, All=AllPrice),
cXnames = "price", data = detergent, n.draws = 500, burnin = 100,
thin = 3, verbose = TRUE)
## summarize the results
summary(res1)
## calculate the quantities of interest for the first 3 observations
pre1 <- predict(res1, newdata = detergent[1:3,])

## load the Japanese election data
data(japan)
## run the multinomial probit model with ordered preferences
res2 <- mnp(cbind(LDP, NFP, SKG, JCP) ~ gender + education + age, data = japan,
verbose = TRUE)
## summarize the results
summary(res2)
## calculate the predicted probabilities for the 10th observation
## averaging over 100 additional Monte Carlo draws given each of MCMC draw.
pre2 <- predict(res2, newdata = japan[10,], type = "prob", n.draws = 100,
verbose = TRUE)

predict.mnp

Posterior Prediction under the Bayesian Multinomial Probit Models

Description

Obtains posterior predictions under a fitted (Bayesian) multinomial probit model. predict method
for class mnp.

Usage

## S3 method for class 'mnp'
predict(object, newdata = NULL, newdraw = NULL, n.draws = 1,
type = c("prob", "choice", "order"), verbose = FALSE, ...)
predict.mnp

Arguments

object An output object from mnp.
newdata An optional data frame containing the values of the predictor variables. Predictions for multiple values of the predictor variables can be made simultaneously if newdata has multiple rows. The default is the original data frame used for fitting the model.
newdraw An optional matrix of MCMC draws to be used for posterior predictions. The default is the original MCMC draws stored in object.
n.draws The number of additional Monte Carlo draws given each MCMC draw of coefficients and covariance matrix. The specified number of latent variables will be sampled from the multivariate normal distribution, and the quantities of interest will be calculated by averaging over these draws. This will be particularly useful calculating the uncertainty of predicted probabilities. The default is 1.
type The type of posterior predictions required. There are four options: type = "prob" returns the predictive probabilities of being the most preferred choice among the choice set, type = "choice" returns the Monte Carlo sample of the most preferred choice, and type = "order" returns the Monte Carlo sample of the ordered preferences,
verbose logical. If TRUE, helpful messages along with a progress report on the Monte Carlo sampling from the posterior predictive distributions are printed on the screen. The default is FALSE.
... additional arguments passed to other methods.

Details

The posterior predictive values are computed using the Monte Carlo sample stored in the mnp output (or other sample if newdraw is specified). Given each Monte Carlo sample of the parameters and each vector of predictor variables, we sample the vector-valued latent variable from the appropriate multivariate Normal distribution. Then, using the sampled predictive values of the latent variable, we construct the most preferred choice as well as the ordered preferences. Averaging over the Monte Carlo sample of the preferred choice, we obtain the predictive probabilities of each choice being most preferred given the values of the predictor variables. Since the predictive values are computed via Monte Carlo simulations, each run may produce somewhat different values. The computation may be slow if predictions with many values of the predictor variables are required and/or if a large Monte Carlo sample of the model parameters is used. In either case, setting verbose = TRUE may be helpful in monitoring the progress of the code.

Value

predict.mnp yields a list of class predict.mnp containing at least one of the following elements:

- A three dimensional array of the Monte Carlo sample from the posterior predictive distribution of the ordered preferences. The first dimension corresponds to the rows of newdata (or the original data set if newdata is left unspecified), the second dimension corresponds to the alternatives in the choice set, and the third dimension indexes the Monte Carlo sample. If n.draws is greater than 1, then each entry will be an average over these additional draws.
A two or three dimensional array of the posterior predictive probabilities for each alternative in the choice set being most preferred. The first dimension corresponds to the rows of newdata (or the original data set if newdata is left unspecified), the second dimension corresponds to the alternatives in the choice set, and the third dimension (if it exists) indexes the Monte Carlo sample. If n.draws is greater than 1, then the third dimension exists and indexes the Monte Carlo sample.

A matrix of the Monte Carlo sample from the posterior predictive distribution of the most preferred choice. The first dimension corresponds to the rows of newdata (or the original data set if newdata is left unspecified), and the second dimension indexes the Monte Carlo sample. n.draws will be set to 1 when computing this quantity of interest.

A matrix of covariates used for prediction

Author(s)

Kosuke Imai, Department of Politics, Princeton University <kimai@princeton.edu>

See Also

mnp; MNP home page at http://imai.princeton.edu/research/MNP.html

summary.mnp

| summary.mnp | Summarizing the results for the Multinomial Probit Models |

Description

summary method for class mnp.

Usage

```r
## S3 method for class 'mnp'
summary(object, CI=c(2.5, 97.5), ...)

## S3 method for class 'summary.mnp'
print(x, digits = max(3,getOption("digits") - 3), ...)
```

Arguments

- `object`: An output object from mnp.
- `CI`: A 2 dimensional vector of lower and upper bounds for the credible intervals used to summarize the results. The default is the equal tail 95 percent credible interval.
- `x`: An object of class summary.mnp.
- `digits`: the number of significant digits to use when printing.
- `...`: further arguments passed to or from other methods.
Value

The object returned by \texttt{summary.mnp} contains the following elements:

- \texttt{call}: The call from \texttt{mnp}.
- \texttt{n.alt}: The total number of alternatives.
- \texttt{base}: The base category used for fitting.
- \texttt{n.obs}: The number of observations.
- \texttt{n.param}: The number of estimated parameters.
- \texttt{n.draws}: The number of Gibbs draws used for the summary.
- \texttt{coef.table}: The summary of the posterior distribution of the coefficients.
- \texttt{cov.table}: The summary of the posterior distribution of the covariance matrix.

This object can be printed by \texttt{print.summary.mnp}.

Author(s)

Kosuke Imai, Department of Politics, Princeton University <kimai@princeton.edu>

See Also

\texttt{mnp}; MNP home page at \url{http://imai.princeton.edu/research/MNP.html}
Index

*Topic datasets
detergent, 3
japan, 4
*Topic methods
coeff.mnp, 2
cov.mnp, 3
predict.mnp, 8
summary.mnp, 10
*Topic models
mnp, 5
coef.mnp, 2
coefficients.mnp(coef.mnp), 2
cov.mnp, 3
detergent, 3
japan, 4
MNP (mnp), 5
mnp, 5
predict.mnp, 8
print.summary.mnp(summary.mnp), 10
summary.mnp, 10