Package 'Rchoice'

July 2, 2014

Title Discrete Choice (Binary, Poisson and Ordered) Models with Random Parameters

Version 0.1

Author Mauricio Sarrias <msarrias86@gmail.com>

Maintainer Mauricio Sarrias <msarrias86@gmail.com>

Description Estimate binary, ordered and count models with random parameters.

Depends R (>= 2.10), maxLik, Formula

Imports sandwich, ggplot2, plotrix, car, lmtest

License GPL (>= 2)

LazyData no

NeedsCompilation no

Repository CRAN

Date/Publication 2014-04-02 09:08:22

R topics documented:

AIC.Rchoice *Akaike's Information Criterion*

Description

Calculate Akaike's information Criterion (AIC) or the Bayesian information Criterion (BIC) for a model of object of class.

Usage

```
## S3 method for class 'Rchoice'
AIC(object, ..., k = 2)

## S3 method for class 'Rchoice'
BIC(object, ...)
```

Arguments

object	a fitted model of class Rchoice
...	additional arguments to be passed to or from other functions
k	a numeric value, use as penalty coefficient for number of parameters in the fitted model

Value

a numeric value with the corresponding AIC or BIC value.

See Also

Rchoice

Examples

```
## Probit model
data("Workmroz")
probit <- Rchoice(lfp ~ k5 + k618 + age + wc + hc + lwg + inc,
                  data = Workmroz , family = binomial('probit'))
summary(probit)

AIC(probit)
BIC(probit)
```

`Articles` *Doctoral Publications*

Description

Data from research by Long(1990) that analizes the scientist's level of publications.

Usage

```
data(Articles)
```

Format

A data frame with 915 observations on the following 6 variables.

`art` Articles during last 3 years of Ph.D.

`fem` 1 if female scientist; else 0

`mar` 1 if married; else 0

`kid5` Number of children 5 or younger

`phd` Prestige of Ph.D. department

`ment` Articles by mentor during last 3 years

Source

- Long, J. S. (1990). The origins of sex differences in science. Social Forces, 68(4), 1297-1316.
- Long, J. S. (1997). Regression models for categorical and limited dependent variables (Vol. 7). Sage.
- Long, J. S., & Freese, J. (2006). Regression models for categorical and limited dependent variables using Stata. Stata Press, College Station, TX.

Examples

```
data(Articles)
```

`Attitudes` *Attituded toward working mothers*

Description

In 1997 and 1989, the General Social Survey asked respondents to evaluate the following statement: "A working mother can establish just as warm and secure a relationship with her children as a mother who does not work".

Usage

```
data(Attitudes)
```

Format

A data frame with 2293 observations on the following 10 variables.

`warm` 1 = Strongly disagree, 2 = disagree, 3 = agree, 4 = strongly agree

`yr89` survey year: 1 = 1989; 0 = 1977

`male` 1 = male; 0 = female

`white` 1 = white; 0 = nonwhite

`age` age in years

`ed` years of education

`prst` occupational prestige

Source

- Clogg, C. C., & Shihadeh, E. S. (1994). Statistical models for ordinal variables. Thousand Oaks, CA: Sage Publications.
- Long, J. S. (1997). Regression models for categorical and limited dependent variables (Vol. 7). Sage.
- Long, J. S., & Freese, J. (2006). Regression models for categorical and limited dependent variables using Stata. Stata Press, College Station, TX.

Examples

```
data(Attitudes)
```

bread.Rchoice *Bread for sandwiches*

Description

Computes the bread of the sandwich covariance matrix

Usage

```
## S3 method for class 'Rchoice'
bread(x, ...)
```

Arguments

x	a fitted model of class Rchoice
...	Other arguments when bread is applied to another class object

Value

the covariance matrix times observations

References

Zeileis A (2006), Object-oriented Computation of Sandwich Estimators. Journal of Statistical Software, 16(9), 1–16.

Examples

```
## Probit model
data("Workmroz")
probit <- Rchoice(lfp ~ k5 + k618 + age + wc + hc + lwg + inc,
                  data = Workmroz , family = binomial('probit'))
summary(probit)

library(sandwich)
bread(probit)
```

cov.Rchoice *Covariance and Correlation matrix of random parameters*

Description

Computes the Variance-Covariance matrix and the Correlation matrix of the random parameters

Usage

```
cov.Rchoice(x)

cor.Rchoice(x)
```

Arguments

x	a object of class Rchoice,
...	further arguments

Value

a matrix with the variance of the random parameters if model is fitted with random coefficients or the correlation matrix if argument `correlation = TRUE` in the fitted model.

References

- Greene, W. H. (2003). Econometric analysis. Pearson Education India.
- Train, K. (2009). Discrete choice methods with simulation. Cambridge university press.

See Also

Rchoice

---

estfun.Rchoice *Gradient for observations*

Description

It extracts the gradient for each observations evaluated at the estimated parameters

Usage

```
## S3 method for class 'Rchoice'
estfun(x, ...)
```

Arguments

x	a fitted model of class Rchoice
...	Other arguments when estfun is applied to another class object

Value

the gradient matrix of dimension n times k

References

Zeileis A (2006), Object-oriented Computation of Sandwich Estimators. Journal of Statistical Software, 16(9), 1–16.

Examples

```
## Probit model
data("Workmroz")
probit <- Rchoice(lfp ~ k5 + k618 + age + wc + hc + lwg + inc,
                  data = Workmroz , family = binomial('probit'))
summary(probit)

library(sandwich)
estfun(probit)
```

`Health` *German Health Care Data*

Description

German Health Care Data, unbalanced panel.

Usage

```
data(Health)
```

Format

A data frame with 27326 observations on the following 27 variables.

`id` person identification number

`female` female =1, male =0

`year` calendar year of the observation

`age` age in years

`hsat` health satisfaction, 0 (low),...,10 (high)

`handdum` handicapped = 1, 0 otherwise

`handper` degree of handicap in percent; 0,100

`hhinc` household nomial monthly net income in German marks/10,000

`hhkids` children under age 16 in the household = 1; otherwise = 0

`educ` years of schooling

`married` married =1, otherwise = 0

`haupts` highest schooling degree is Hauptschul degree = 1; otherwise = 0

`reals` highest schooling degree is Realschul degree = 1, otherwise = 0

`fachhs` highest schooling degree is Polytechical degree = 1; otherwise = 0

`abitur` highest schooling degree is Abitur = 1; otherwise = 0

`univ` highest schooling degree is university degree =1; otherwise = 0

`working` employed =1; otherwise = 0

`bluec` blue-collar employee = 1; otherwise = 0

`whitec` white-collar employeee =1; otherwise = 0

`self` self-employed = 1; otherwise = 0

`beamt` civil servant = 1; otherwise = 0

`docvis` number of doctor visits in last three months

`hospvis` number of hospital visits in last calendar year

`public` insured in public health =1; otherwise = 0

`addon` insured by add-on insurance =1; otherwise = 0

`hsat2` 40 observations on hsat recorded between 6 and 7 were changed to 7

`newhsat` recording of hsat, (0-2) = 0, (3-5)=1, (6-8)=2, (9)=3 (10)=4

Source

Riphahn, R. T., Wambach, A., & Million, A. (2003). Incentive effects in the demand for health care: a bivariate panel count data estimation. Journal of applied econometrics, 18(4), 387-405.

References

Greene, W. H. (2003). Econometric analysis. Pearson Education India.

Examples

```
data(Health)
```

---

plot.Rchoice *Plot random parameters*

Description

Plot the conditional expectation of random parameters estimated by `Rchoice`.

Usage

```
## S3 method for class 'Rchoice'
plot(x, par = NULL, ind = FALSE, id = NULL,
  type = c("density", "histogram"), bin = 1, adjust = 1, ...)
```

Arguments

x	a object of class Rchoice,
par	a string giving the name of the variable with random parameter,
type	a string indicating the type of distribution: it can be a histogram or a density of the conditional expectation of the random coefficients,
ind	a boolean. If TRUE, a 95 conditional distribution for each individual is plotted. As default, the conditional expectation of par for the first 10 individual is plotted,
id	only relevant if ind is not NULL. This is a vector indicating the position of the individual for whom the user want to plot the conditional coefficients,
bin	bin of histrogram,
adjust	bandwidth for the kernel density,
...	further arguments to be passed to qplot or plotCI,

Value

a plot with the distribution or a confident interval of the conditional random coefficients.

References

- Greene, W. H. (2003). Econometric analysis. Pearson Education India.
- Train, K. (2009). Discrete choice methods with simulation. Cambridge university press.

See Also

Rchoice, ggplot2

Rchoice *Estimate discrete choice model with random parameters*

Description

Estimation of discrete choice models such as Binary (logit and probit), Poisson and Ordered (logit and probit) model with random coefficients for cross-section data by simulated maximum likelihood

Usage

```
Rchoice(formula, data, subset, weights, na.action, family, start = NULL,
  ranp = NULL, R = 40, haltons = NA, seed = 10, correlation = FALSE,
  ...)

## S3 method for class 'Rchoice'
terms(x, ...)

## S3 method for class 'Rchoice'
```

```
model.matrix(object, ...)

## S3 method for class 'Rchoice'
vcov(object, ...)

## S3 method for class 'Rchoice'
coef(object, ...)

## S3 method for class 'Rchoice'
nObs(x, ...)

## S3 method for class 'Rchoice'
fitted(object, ...)

## S3 method for class 'Rchoice'
df.residual(object, ...)

## S3 method for class 'Rchoice'
update(object, new, ...)

## S3 method for class 'Rchoice'
logLik(object, ...)

## S3 method for class 'Rchoice'
print(x, digits = max(3, getOption("digits") - 3),
  width = getOption("width"), ...)

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

getSummary.Rchoice(obj, alpha = 0.05, ...)
```

Arguments

x,object,obj	and object of class Rchoice,
formula	a symbolic description of the model to be estimated,
new	an updated formula for the update method,
data	the data,
subset	an optional vector specifying a subset of observations,
weights	an optional vector of weigths,
na.action	a function wich indicated what should happen when the data contains NA's
start	a vector of starting values,
family	the distribution to be used,
ranp	a named vector whose names are the random parameters and values the distribution: "n" for normal, "ln" for log-normal, "cn" for truncated normal, "u" for uniform, "t" for triangular,

R	the number of draws of pseudo-random numbers if ranp is not NULL.
haltons	only relevant if ranp is not NULL. If not NULL, halton sequence is used instead of pseudo-random numbers. If haltons=NA, some default values are used for the prime of the sequence and for the number of element droped. Otherwise, haltons should be a list with elements prime and drop.
seed	,
correlation	only relevant if ranp is not NULL. If true, the correlation between random parameters is taken into account,
alpha	significance value for getSummary,
digits	number of digits,
width	width,
...	further arguments passed to maxLik.

Details

The models are estimated using the maxLik function of maxLik package.

If ranp is not NULL, the random parameter (random coefficient) model is estimated. A random parameter model or random coefficient models permits regression parameter to vary across individuals according to some distribution. A fully parametric random parameter model specifies the latent variable y^* conditional on regressors x and given parameters β_i to have conditional density $f(y|x, \beta_i)$ where β_i are iid with density $g(\beta_i|\theta_i)$. The density is assumed a priori by the user by the argument ranp. If the parameters are assumed to be normally distributed $\beta_i \tilde{} N(\beta, \Sigma)$, then the random parameter are constructed as:

$$\beta_{ir} = \beta + L\omega_{ir}$$

where $LL' = \Sigma$ and ω_{ir} is the r-th draw from standard normal distribution for individual i.

Once the model is specified by the argument family, the model is estimated using Simulated Maximum Likelihood (SMLE). The probabilities, given by $f(y|x, \beta_i)$, are simulated using R pseudo-draws if halton=NULL or R halton draws if halton = NA. The user can also specified the primes and the number of dropped elements for the halton draws. For example, if the model consists of two random parameters, the user can specify haltons = list("prime" = c(2,3), "drop" = c(11, 11)).

A random parameter hierarchical model can be estimated by including heterogeneity in the mean of the random parameters:

$$\beta_{ir} = \beta + \pi' s_i + L\omega_{ir}$$

Rchoice manages the variables in the hierarchical model by the formula object: all the hierarchical variables (s_i) are included after the | symbol. See examples below

Value

An object of class "Rchoice", a list elements:

coefficients	the named vector of coefficients,
family	type of model,
link	distribution of the errors,

logLik	a set of values of the maximum likelihood procedure,
mf	the model framed used,
formula	the formula (a Formula object),
time	proc.time() minus the start time,
freq	frequency of dependent variable,
draws	type of draws used,
R.model	TRUE if a random parameter model is fitted,
R	number of draws used,
b.random	matrix of conditional expectation of random parameters,
sd.random	matrix of standard deviation of conditional expectation of random parameters,
ranp	vector indicating the variables with random parameters and their distribution,
probabilities	the fitted probabilities for each individuals,
residuals	the residuals,
call	the matched call.

Author(s)

Mauricio Sarrias <msarrias86@gmail.com>

References

Greene, W. H. (2012). Econometric analysis. 7 edition. Prentice Hall.

Train, K. (2009). Discretechoice methods with simulation. Cambridge university press.

See Also

mlogit, maxLik

Examples

```
## Probit model
data("Workmroz")
probit <- Rchoice(lfp ~ k5 + k618 + age + wc + hc + lwg + inc,
                  data = Workmroz, family = binomial('probit'))
summary(probit)

## Poisson model
data("Articles")
poisson <- Rchoice(art ~ fem + mar + kid5 + phd + ment, data = Articles, family = poisson)
summary(poisson)

## Ordered probit model
data("Health")
oprobit <- Rchoice(newhsat ~ age + educ + hhinc + married + hhkids,
data = Health, family = ordinal('probit'), subset = year == 1988)
summary(oprobit)
```

```
## Not run:
## Hierarchical Logit Random Parameter Model
Hran.logit<-Rchoice(lfp ~ k618 + lwg + wc + inc + k5 | age + wc + hc,
ranp = c(inc = "t", k5 = "n"),
family = binomial('logit'), data = Workmroz)
summary(Hran.logit)

## End(Not run)

## Not run:
## Hierarchical Poisson model with correlated random parameters
poissonH.ran <- Rchoice(art ~ fem + mar + kid5 + phd + ment | fem, data = Articles,
ranp = c(kid5 = "n", phd = "n", ment = "n"), family = poisson, correlation =  TRUE)
summary(poissonH.ran)

## End(Not run)

## Not run:
## Ordered Probit model with random parameters
oprobit.ran <- Rchoice(newhsat ~ age + educ + hhinc + married + hhkids,
                       data = Health, family = ordinal('probit'),
                       subset = year == 1988,
                       ranp = c(age = "n", hhinc = "n"),
                       start = rep(0, 11))
summary(oprobit.ran)

## End(Not run)
```

`Workmroz` *Labor Force Participation*

Description

Data extracted by Mroz(1987) from the 197 Panel Study of Income Dynacmis. The sample consists of 753 white, married women between the ages of 30 and 60.

Usage

```
data(Workmroz)
```

Format

A data frame with 753 observations on the following 9 variables.

`lfp` 1 if wife is in the paid labor force; else 0

`k5` Number of children ages 5 and younger

`k618` Number of children ages 6 to 18

`age` Wife's age in years

`wc` 1 if wife attended college; else 0

`hc` 1 if husband attended college; else 0

`lwg` Log of wife's estimated wage rate

`inc` Family income excluding wife's wage

`linc` Log of Family income excluding wife's wage

Source

Mroz, T. A. (1987). The sensitivity of an empirical model of married women's hours of work to economic and statistical assumptions. Econometrica, 55(4), 765-799

Examples

```
data(Workmroz)
```

Index