Package ‘abms’

April 5, 2025

Title Augmented Bayesian Model Selection for Regression Models

Version 0.2

Description Tools to perform model selection alongside estimation under Linear, Logistic, Negative binomial, Quantile, and Skew-Normal regression. Under the spike-and-slab method, a probability for each possible model is estimated with the posterior mean, credibility interval, and standard deviation of coefficients and parameters under the most probable model.

License GPL (>= 3)

URL https://github.com/SirCornflake/BMS

Encoding UTF-8

RoxygenNote 7.2.2

Imports BayesLogit, GIGrvg, mvtnorm, truncnorm

Depends R (>= 3.6.0)

LazyData true

NeedsCompilation no

Author Francisco Segovia [aut, cre],
Luis Gutierres [aut],
Ramses Mena [aut]

Maintainer Francisco Segovia <fasegovia@uc.cl>

Repository CRAN

Date/Publication 2025-04-05 21:30:03 UTC

Contents

ens 2
gen_base_binomial_reg 3
gen_base_NegBinomial_reg 4
gibbs_abms 5
rCRT 8
summary_gibbs 9
womack 10

Index **11**

ens *Chilean National Health Survey (2016-2017)*

Description

`ens` is a data-base that reo collected by the Chilean National Health Survey in order to study the health status of the population and policy making. This survey was performed between 2016 and 2017

Usage

```
ens
```

Format

A Data frame with 3238 rows and 15 variables (columns):

`pas` Blood systolic pressure

`pad` Blood diastolic pressure

`age` In years

`waist` diameter in centimeters

`bmi` body mass index

`weight` In centimeters

`height` In centimeters

`hypertension` Presence of hypertension (1: Yes, 0: No)

`sedentary` sedentary person (1: Yes, 0: No)

`smoker` 1: Yes, 0: No

`diabetes` 1: Yes, 0: No

`depression` 1: Yes, 0: No

`male` 1: Yes, 0: female

`scholar` Years of formal education

Source

Chilean Health Ministry (https://epi.minsal.cl/encuesta-ens-anteriores/)

Examples

```
data(ens)
```

gen_base_binomial_reg *Logistic Regression Data generator*

Description

It generates N observations of the Binomial distribution with parameters ni (the i-th's individual sizes) and p (the success probability), where the coefficients are indexed on p via the logistic function.

Usage

```
gen_base_binomial_reg(N, beta, Covariates, ni = rep(1, N))
```

Arguments

N	The number of observations that will be generated. It must be a positive integer.
beta	A vector of coefficients including the intercept. It can be a matrix.
Covariates	A data.frame object with the predictors (without intercept) for which we want to test if they are relevant to the response variable. It can also be a (nxp) matrix.
ni	A vector of size n that represent the i-th individual size (the size parameter of the binomial distribution). It can also be a $(nx1)$ matrix. For default, all individual size are fixed at 1.

Value

The function return a table with the sample of size N from the Binomial distribution indexed with the predictors indicated in the Covariates argument, the ni, the number of failures (ni - y), and the predictors for each individual .

Examples

```
N<-200    #Number of extractions
beta<-c(1, 0, 2, 0, 3, 2)    #Coefficient vector
p<-length(beta)
aux_cov<-rnorm((p-1)*N, 0,1)
Covariates<-data.frame(matrix(aux_cov, ncol=p-1, nrow=N))  #Generating the Covariates data.frame
colnames(Covariates)<-c("X1", "X2", "X3", "X4", "X5")
base<-gen_base_binomial_reg(N, beta, Covariates, ni=rep(1, N))    #Generating the data
base
```

gen_base_NegBinomial_reg
Negative Binomial Regression Data generator

Description

It generates N observations of the Negative binomial distribution with parameters r (number of success) and p (success probability), where the coefficients are indexed on p via the logistic function.

Usage

```
gen_base_NegBinomial_reg(N, beta, r, Covariates)
```

Arguments

N	The number of observations that will be generated. It must be a positive integer.
beta	A vector of coefficients including the intercept. It can be a matrix.
r	The number of success parameter. It must be a positive integer.
Covariates	A data.frame object with the predictors (without intercept) for which we want to test if they are relevant to the response variable. It can also be a (nxp) matrix.

Value

The function return a sample of size N from the Negative binomial distribution indexed with the predictors indicated in the Covariates argument, and the predictors for each individual.

Examples

```
N<-10   #Number of extractions
beta<-c(0.5, -0.8,  1.0,  0,  0.4, -0.7)  #Coefficient vector
p<-length(beta)
r<-2    #Number of success parameter
aux_cov<-rnorm((p-1)*N, 0,1)
Covariates<-data.frame(matrix(aux_cov, ncol=p-1, nrow=N))  #Generating the Covariates data.frame
colnames(Covariates)<-c("X1", "X2", "X3", "X4", "X5")
base<-gen_base_NegBinomial_reg(N, beta, r, Covariates)    #Generating the data
base
```

`gibbs_abms` *Bayesian variable selection models via a spike-and-slab methodology.*

Description

A Bayesian model selection methodology based on the spike-and-slab strategy and an augmentation technique for Linear, Logistic, Negative Binomial, Quantile, and Skew Normal Regression. The model considers a response vector y of size n and p predictors to perform coefficient estimation and asses which ones are relevant to explain the response distribution. Other parameters related to the family selected are also estimated. Summary results can be provided using the `summary_gibbs()` R function.

Usage

```
gibbs_abms(
  y,
  Covariates,
  family = "LiR",
  first_excluded = 0,
  nchain = 10000,
  burnin = 2000,
  tau2 = 1000,
  rho = 1,
  ni = rep(1, length(y)),
  alpha = 0.5,
  a0 = 1,
  b0 = 1,
  d = 2,
  b2 = 1/2,
  model_fixed = NULL,
  WomackPrior = TRUE,
  a_bb = 1,
  b_bb = 1,
  count.iteration = TRUE
)
```

Arguments

`y` A vector of size n with observed responses. It can also be a $(nx1)$ matrix.

`Covariates` A data.frame object with the predictors (without the intercept) for which we want to test if they are relevant to the response variable. It can also be a (nxp) matrix.

`family` A character object that describes the hierarchical regression model that will be used. If `family="LiR"`, then a Linear regresion model will be fitted (gaussian errors). If `family="LoR"`, then a Logistic regresion model will be fitted (binomial distribution). If `family="NBR"`, then a Negative Binomial regresion model will be fitted (mean $r(1-p)/p$). If `family="QR"`, then a Quantile regresion model will be fitted (Asymmetric Laplace errors). If `family="SNR"`, then a Skew normal regresion model will be fitted (Skew-Normal errors). The argument is fixed at `family="LiR"` by default.

`first_excluded` A non-negative integer that indicates which first columns will not be tested. For example, if `first_excluded=2`, the two first columns of `Covariates` will not be tested. Intercept is always excluded for the selection process.

`nchain` The Gibbs sampler's chain size, it must be a non-negative integer. The default value is 10,000

`burnin` The burn-in period of the Gibbs sampler, it must be a non-negative integer and greater than `nchain`. The default value is 2,000

`tau2` The variance prior of each coefficient, it must be a positive real number. Fixed at 1 by deafault

`rho` The parameter of the Womack prior, it must be a positive real number. Fixed at 1 by deafault

`ni` For Logistic regression only. A vector of size n that represent the i-th individual size (the size parameter of the Binomial distribution) that it must be a positive integer. It can also be a $(nx1)$ matrix. For default, all individual size are fixed at 1.

`alpha` For Quantile regression only. The desired quantile for which we want to perform Quantile regression. `alpha` must be between $(0, 1)$. By default, `alpha=0.5`, that is, median regression.

`a0` This argument depends on the family choosen. For `family="LiR"`, is the shape hyper-parameter of the $Gamma$ prior to the variance parameter (σ^2) of the Gaussian distribution. For `family="NBR"` is the shape hyper-parameter of the $Gamma$ prior to the parameter r the Negative Binomial distribution (the number of successes until the experiment is stopped). For `family="QR"` is the shape hyper-parameter of the $Gamma$ prior to thevariance parameter (σ^2) of the Asymmetric Laplace distribution. Note thas this argument do not exist for `family=LoR` and `family=SNR`. For all hierarchical regression models, it must be a positive real number and its fixed at 1 by deafault.

`b0` This argument depends on the family choosen. For `family="LiR"` is the scale hyper-parameter of the $Gamma$ prior to the variance parameter (σ^2) of the Gaussian distribution. For `family="NBR"` is the scale hyper-parameter of the $Gamma$ prior to the parameter r the Negative Binomial distribution (the number of successes until the experiment is stopped). For `family="QR"` is the scale hyper-parameter of the $Gamma$ prior to the variance parameter (σ^2) of the Asymmetric Laplace distribution. Note thas this argument do not exist for `family=LoR` and `family=SNR`. For all hierarchical regression models, it must be a positive real number and its fixed at 1 by deafault.

`d` For the Skew-Normal regression only. It is the location hyper-parameter of the t-student prior to the parameter $lambda$ (asymmetric parameter of the Skew-Normal distribution). By default is fixed at 2, which is recommended.

`b2` For the Skew-Normal regression only. It is the scale hyper-parameter of the t-student prior to the parameter lambda (asymmetric parameter of the Skew-Normal distribution). By default is fixed at 1/2, which is recommended.

model_fixed
: Either NULL or a vector that indicates which model will be fixed to perform parameter estimation only under such a model. For example, if there are only three predictors and model.fixed=c(1,3), only parameter estimation will be performed where only the first and third predictors are included. If NULL, model selection will also be performed. Fixed at NULL by default.

WomackPrior
: A logical argument. If TRUE, the Womack prior for the model space will be used. Otherwise, the Beta-Binomial prior with shape parameters a_bb and b_bb will be used. Fixed at TRUE by default

a_bb
: A numeric vector of length 1. The first shape parameter of the Beta-Binomial prior. Recomended value is a_bb=1.

b_bb
: A numeric vector of length 1. The second shape parameter of the Beta-Binomial prior. Recomended value is b_bb=p_selection^u, where u>1, and p_selection is the number of predictors under the selection process.

count.iteration
: A logical argument. If TRUE, a counter for the Gibbs sampler iterations will be displayed. Fixed at TRUE by deafult.

Value

A abms object with the following variables:

family
: This character object prints the name of the fitted hierarchical regression model. It needs to be extracted from the list 'Default'.

prednames
: A character object that prints the predictors names, using the columns names of the Covariates argument. It needs to be extracted from the list 'Default'.

Seconds
: How many seconds the method took. It needs to be extracted from the list 'Default'.

tau2
: The tau2 that was used as argument.

y
: The y response vector that was used as argument.

Covariates
: The Covariates data frame or matrix that was used as argument.

beta_chain
: The coefficients sample for each Gibbs sampler iteration. A (nchain x p) matrix

sigma2_chain
: For the Linear, Quantile and Skew-Normal regression only. The variance parameter (σ^2) sample for each Gibbs sampler iteration. A (nchain x 1) matrix

r_chain
: For the Negative-Binomial regression only. The number of failure parameter (r) sample for each Gibbs sampler iteration. A (nchain x 1) matrix

lambda_chain
: For the Skew-Normal regression only. The asymmetric parameter (λ) sample for each Gibbs sampler iteration. A (nchain x 1) matrix

model_chain
: The model selected at each Gibbs sampler iteration. A (nchain x p) matrix.

Z_chain
: For internal use.

t_chain
: For internal use.

References

Azzalini (1985). A class of distributions which includes the normal ones, Scandinavian Journal of Statistics 12(2): 171:178.

Bayes, C. and Branco, M. (2007). Bayesian inference for the skewness parameter of the scalar skew-normal distribution. Brazilian Journal of Probability and Statistics. 21: 141:163.

Kotz, S., Kozubowski, T. and Podgorski, K. (2001). The Laplace Distribution and Generalization, first edn, Birkhauser Basel.

Polson, N., Scott, J., and Windle, J. (2013). Bayesian Inference for Logistic Models Using Polya Gamma Latent Variables. Journal of the American Statistical Association, 108: 1339:1349.

Zhou, W. and Carin, L. (2013). Negative Binomial Process Count and Mixture Modeling. arXiv:1405.0506v1.

Examples

```
##################################################
##      Gibbs for Linear Regression  ##
##################################################

## Simulating data
set.seed(31415)
N<-200
r_beta<-as.matrix(c(1, 0, 2, 0))
r_p<-length(r_beta)
r_sigma2<-1.5
X<-matrix( c(rep(1, N), rnorm((r_p -1)*N)), ncol=r_p )
Xbeta<-X%*%r_beta
y<-rnorm(N, mean=Xbeta , sd=sqrt(r_sigma2))
Covariates<-X[,2:(length(r_beta))];
colnames(Covariates)<-c("X1", "X2", "X3")

## Fitting the model
fit<- gibbs_abms(y, Covariates, family="LiR", first_excluded=0, nchain=1000, burnin=20,
 a0=1, b0=1)

summary_gibbs(fit, BF=FALSE) #Summary results

## For more examples, see "Model Ilustrations.R" file in
## https://github.com/SirCornflake/BMS
```

---

rCRT *Title*

Description

Sampling from the Chines Restaurant distribution.

Usage

```
rCRT(n, b, c)
```

Arguments

n	Number of observations. It must be a positive integer.
b	Parameter distribution, a non-negative integer. The number of Bernoulli independent variables that are added. It can be a vector
c	Parameter distribution, a positive real number. Used calculate the success probability the j-th Bernoulli independent variable, that is, c/(c +j -1). It can be a vector

Value

This function generates n random variables from the $CRT(b, c)$ distribution

References

Pitman, Jim (1995). "Exchangeable and Partially Exchangeable Random Partitions". Probability Theory and Related Fields. 102 (2): 145:158

Examples

```
#Generating 4 random variables with parameters b=2 and c=1
rCRT(4,2,1)
```

---

summary_gibbs   *Summary function for abms objects*

Description

For abms objects, it returns the posterior mean, standard deviation, and 95% centered credible interval for each parameter. Additionally, it provides all explored models alongside the conditional Bayes factors and marginal Bayes factors estimator between the most probable model and the others that have arisen.

Usage

```
summary_gibbs(fit, BF = FALSE)
```

Arguments

fit	An abms object. Such object is obtained by fitting a regression model with the `gibbs_abms()` function.
BF	A logical object. if `TRUE`, then the Bayes factor comparison is shown.`BF=FALSE` by default.

Value

A summary of the inference performed by the Bayesian model obtained by the `gibbs_abms()` function. The variables are:

- `Mean_IC` A table with the posterior mean, standard deviation, and 95% centered credible interval for each parameter
- `Explored_Models` A table with all explored models. If `BF=TRUE`, the conditional Bayes factors and marginal Bayes factors estimator between the most probable model and the others that have arisen are displayed.

Examples

```
## See \code{gibbs_abms()} help page function
```

---

womack *For internal use Womack prior*

Description

Womack probability mass function with 'K' predictors and parameter 'rho'.

Usage

```
womack(K, rho)
```

Arguments

- `K` Number of predictors. It must be a positive integer.
- `rho` Value for the "rho" parameter. It must be positive real number

Value

Given that all models of the same hierarchy has the same prior probability, this function returns one value for each hierarchy. including the null model (size=0)

References

Womack, A., Fuentes, C., and Rodriguez-Taylor, D. (2015). "Model Space Priors for Objective Sparse Bayesian Regression." arXiv:1511.04745. 8:24

Examples

```
#Fixing rho=1 and 3 predictors
womack(K=3, rho=1)
#it returns Womack prior for all models of size 0 (the null model), 1,2 and 3 (the full model)
```

Index

* **datasets**
    ens, 2

ens, 2

gen_base_binomial_reg, 3
gen_base_NegBinomial_reg, 4
gibbs_abms, 5

rCRT, 8

summary_gibbs, 9

womack, 10