Package 'BayesGESM'

June 6, 2015

Type Package

Title Bayesian Analysis of Generalized Elliptical Semi-Parametric Models and Flexible Measurement Error Models

Version 1.4

Date 2015-06-04

Author Luz Marina Rondon <lumarp@gmail.com> and Heleno Bolfarine

Maintainer Luz Marina Rondon <lumarp@gmail.com>

Description Set of tools to perform the statistical inference based on the Bayesian approach for regression models under the assumption that independent additive errors follow normal, Student-t, slash, contaminated normal, Laplace or symmetric hyperbolic distributions, i.e., additive errors follow a scale mixtures of normal distributions. The regression models considered in this package are: (i) Generalized elliptical semi-parametric models, where both location and dispersion parameters of the response variable distribution include non-parametric additive components described by using B-splines; and (ii) Flexible measurement error models under the presence of homoscedastic and heteroscedastic random errors, which admit explanatory variables with and without measurement additive errors as well as the presence of a non-parametric components approximated by using B-splines.

License GPL-2 | GPL-3

Depends splines, GIGrvg, normalp, Formula

Suggests ssym

NeedsCompilation no

Repository CRAN

Date/Publication 2015-06-06 07:50:04

R topics documented:

`BayesGESM-package` *Bayesian Analysis of Generalized Elliptical Semi-Parametric Models and Flexible Measurement Error Models*

Description

This package allows to perform the statistical inference based on the Bayesian approach for regression models under the assumption that independent additive errors follow normal, Student-t, slash, contaminated normal, Laplace or symmetric hyperbolic distributions, i.e., additive errors follow a scale mixtures of normal distributions. The regression models considered in this package are: *(i)* Generalized elliptical semi-parametric models, where both location and dispersion parameters of the response variable distribution include non-parametric additive components described by using B-splines; and *(ii)* Flexible measurement error models under the presence of homoscedastic and heteroscedastic random errors, which admit explanatory variables with and without measurement additive errors as well as the presence of non-parametric components approximated by using B-splines.

Details

Package:	BayesGESM
Type:	Package
Version:	1.4
Date:	2015-06-04
License:	GPL-2 | GPL-3

Author(s)

Luz Marina Rondon <lumarp@gmail.com> and Heleno Bolfarine

Maintainer: Luz Marina Rondon

References

Rondon, L.M. and Bolfarine, H. (2015) Bayesian Analysis of Generalized Elliptical Semi-parametric Models. (submitted).

Rondon, L.M. and Bolfarine, H. (2015). Bayesian analysis of flexible measurement error models.(submitted)

Examples

```
######### Example for Generalized Elliptical Semi-parametric Models #####
#library(ssym)
#data(Erabbits)
#Erabbits2 <- Erabbits[order(Erabbits$age,Erabbits$wlens),]
#attach(Erabbits2)

#fit <- gesm(wlens ~ bsp(age) | bsp(age), family= "ContNormal", eta=c(0.8,0.9),
#  burn.in=1000, post.sam.s=5000, thin=2)
#summary(fit)

######### Example for Flexible Measurement Error Models ################
#### Ragweed Pollen ####
#library(SemiPar)
#data(ragweed)
#attach(ragweed)
#ragweedn <- as.data.frame(ragweed[year==1993,])
#
#model <- fmem(sqrt(ragweed) ~ wind.speed | rain + temperature + bsp(day.in.seas),
#    data=ragweedn,family="Normal", burn.in=500, post.sam.s=2000,
#    thin=10, omeg=1)
#summary(model)
# bsp.graph.fmem(model, 1, xlab="day.in.seas", ylab="f(day.in.seas)")
#
#
#### Boston Data set #########
#library(MASS)
#data(Boston)
#attach(Boston)
#model <- fmem(log(medv) ~ nox | crim + rm + bsp(lstat) + bsp(dis), data=Boston,
#              family="ContNormal", burn.in=10000, post.sam.s=5000, omeg=4, thin=10)
#summary(model)
#bsp.graph.fmem(model,1) ### for variable lstat
#bsp.graph.fmem(model,2) ### for variable dis
#
```

---

bsp *Tool to approximate smooth functions by B-splines.*

Description

bsp is used to approximate smooth functions by B-splines.

Usage

```
bsp(x,kn)
```

Arguments

`x`	values of the explanatory variable.
`kn`	(optional) the number of internal knots . The default value is $[n^{1/5}]$, where *n* is the sample size.

Details

This function uses the routine `bs()` of the R package *splines*.

Value

`x`	the B-spline basis matrix, which is cubic.

Author(s)

Luz Marina Rondon <lumarp@gmail.com> and Heleno Bolfarine

References

De Boor, C. (1978). A practical Guide to Splines. Applied Mathematical Sciences. Springer-Verlag, New York.

`bsp.graph.fmem` *Tool for plotting the nonlinear effects that are approximated by using B-splines for flexible measurement error models*

Description

bsp.graph.fmem displays the graphs of the nonparametric effects from an object of the class `fmem()`.

Usage

```
bsp.graph.fmem(object, which, xlab, ylab, main)
```

Arguments

`object`	An object of the class `fmem()`.
`which`	An integer value, that indicates which nonparametric effect is required.
`xlab`	(optional) A title for the *x* axis
`ylab`	(optional) A title for the *y* axis
`main`	(optional) An overall title for the graph.

Author(s)

Luz Marina Rondon <lumarp@gmail.com> and Heleno Bolfarine `bsp.graph.gesm` *Tool for plotting the nonlinear effects that are approximated by using B-splines for generalized elliptical semi-parametric models*

Description

bsp.graph displays the graphs of the nonparametric effects from an object of the class `gesm()`.

Usage

```
bsp.graph.gesm(object, which, var, xlab, ylab, main)
```

Arguments

`object`	An object of the class `gesm()` or `fmem()`.
`which`	An integer value, where 1 indicates location submodel, and 2 indicates dispersion submodel.
`var`	The name of the variable that is approximate using B-splines.
`xlab`	(optional) A title for the x axis
`ylab`	(optional) A title for the y axis
`main`	(optional) An overall title for the graph.

Author(s)

Luz Marina Rondon <lumarp@gmail.com> and Heleno Bolfarine

`fmem` *Flexible Measurement Error Models*

Description

fmem is used to obtain the statistical inference based on the Bayesian approach for the structural version of the flexible measurement error models under the presence of homoscedastic and heteroscedastic random errors. These models admits vectors of explanatory variables with and without measurement error as well as the presence of nonlinear effects, which is approximated by using B-splines. The error-prone variables and the random error follow scale mixtures of normal distributions.

Usage

```
fmem(formula, data, omeg, family, eta, burn.in, post.sam.s, thin, heter)
```

Arguments

`formula` a symbolic description of the systematic component of the model to be fitted. See details for further information.

`data` an optional data frame, list or environment containing the variables in the model.

`omeg` (optional) the ratio $\omega = \sigma^2_\epsilon / \sigma^2_\xi$, this value must be specified when the model of interest is the homocedastic flexible measurement model. If this value is not specified is assumed to be 1, that is, $\sigma^2_y = \sigma^2_\xi$

`family` a description of the error-prone variables and the random error distributions to be used in the model. Supported distributions include *Normal*, *Student-t*, *Slash*,*Hyperbolic*, *Laplace* and *ContNormal*, which correspond to normal, Student-t, slash, symmetric hyperbolic, Laplace and contaminated normal distributions, respectively.

`eta` (optional) a numeric value or numeric vector that represents the extra parameter of the specified error distribution. This parameter can be assumed known or unknown.

`burn.in` the number of burn-in iterations for the MCMC algorithm.

`post.sam.s` the required size for the posterior sample of interest parameters.

`thin` (optional) the thinning interval used in the simulation to obtain the required size for the posterior sample.

`heter` (optional) An object type list that contains the values $\sigma^2_{\epsilon_i}$ and Σ_{ξ_i} for all i ($i = 1, ldots, n$). The objects have to be specified as `sigma2y` and `sigma2xi`, i.e. `heter <- list(sigma2y, sigma2xi)`. If this argument is not specified the adjusted model is the version homocedastic.

Details

The argument *formula* comprises of three parts, namely: *(i)* observed response variable; *(ii)* covariates with measurement error; and *(iii)* covariates without measurement error including the non-parametric components, which can be specified by using the function `bsp()`. The first two parts are separated by the symbol "~" and the second and third parts are separated by the symbol "|".

This function allows to fit the measurement error model under the presence of homocedastic and heterocedastic random errors. These models admits vectors of explanatory variables with and without measurement error as well as the presence of nonlinear effects approximated by using B-splines. The model investigated is the structural version, as the error-prone variables follow scale mixtures of normal distributions.

Value

`chains` A matrix that contains the posterior sample of interest parameters. Each column represents the marginal posterior sample of each parameter.

`res` a vector of quantile residuals, proposed by Dunn and Smyth (1996) in the context of classical inference, but suited here to the Bayesian case.

`K-L` a vector of case-deletion influence measures based on the Kullback-Leibler divergence.

X_2	a vector of case-deletion influence measures based on the X2-Distance divergence.
DIC	DIC criterion for model selection.
LMPL	Log-marginal pseudo-likelihood for model selection.

Author(s)

Luz Marina Rondon <lumarp@gmail.com> and Heleno Bolfarine

References

Rondon, L.M. and Bolfarine, H. (2015). Bayesian analysis of flexible measurement error models. (submitted)

See Also

bsp bsp.graph.fmem

Examples

```
#library(SemiPar)
#### Ragweed Pollen ####
#data(ragweed)
#attach(ragweed)
############ Example ragweed data
#ragweed2 <- ragweed[year==1993]
#day.in.seas  <- day.in.seas[year==1993]
#temperature <- temperature[year==1993]
#rain <- rain[year==1993]
#wind.speed <- wind.speed[year==1993]
#ragweedn <- data.frame(ragweed2,day.in.seas,temperature,rain,wind.speed)

#model <- fmem(sqrt(ragweed2) ~ wind.speed | rain + temperature + bsp(day.in.seas),
#    data=ragweedn,family="Normal", burn.in=500, post.sam.s=2000,
#    thin=10, omeg=1)
#summary(model)
#
###   Plot non-parametric component
#bsp.graph.fmem(model, which=1, xlab="Day", ylab="f(Day)")


############ Example Boston data
#library(MASS)
#data(Boston)
#attach(Boston)

#model <- fmem(log(medv) ~ nox | crim + rm + bsp(lstat) + bsp(dis), data=Boston,
#               family="ContNormal", burn.in=10000, post.sam.s=5000, omeg=4, thin=10)
#summary(model)
#
###   Plot non-parametric components
```

```
#bsp.graph.fmem(model, which=1, xlab="lstat", ylab="f(lstat)") ### for variable lstat
#bsp.graph.fmem(model, which=2, xlab="dis", ylab="f(dis)") ### for variable dis
#
```

---

gesm *Generalized Elliptical Semi-parametric Models*

Description

gesm is used to obtain the statistical inference based on the Bayesian approach for regression models under the assumption that independent additive errors follow a scale mixtures of normal distribution (i.e., normal, Student-t, slash, contaminated normal, Laplace and symmetric hyperbolic distribution), where both location and dispersion parameters of the response variable distribution include nonparametric additive components described by B-splines.

Usage

```
gesm(formula, data, family, eta, burn.in, post.sam.s, thin)
```

Arguments

`formula`	a symbolic description of the systematic component of the model to be fitted. This description allows parametric and nonparametric functions in the locacion and dispersion parameters. See details for further information.
`data`	an optional data frame, list or environment containing the variables in the model.
`family`	a description of the error distribution to be used in the model. Supported distributions include *Normal*, *Student-t*, *Slash*, *Hyperbolic*, *Laplace* and *ContNormal*, which correspond to normal, Student-t, slash, symmetric hyperbolic, Laplace and contaminated normal distributions, respectively.
`eta`	(optional) a numeric value or numeric vector that represents the extra parameter of the specified error distribution. This parameter can be assumed known or unknown.
`burn.in`	the number of burn-in iterations for the MCMC algorithm.
`post.sam.s`	the required size for the posterior sample of interest parameters.
`thin`	(optional) the thinning interval used in the simulation to obtain the required size for the posterior sample.

Details

The argument *formula* comprises three parts, namely: *(i)* observed response variable; *(ii)* covariates for the location parameter including the nonparametric components; and *(iii)* covariates for the dispersion parameter including the nonparametric components. The first two parts are separated by the symbol "~" and the second and third parts are separated by the symbol "|". Furthemore, the nonparametric components can be specified by using the function `bsp()` in the second and third parts of the argument *formula*.

We implemented an efficient MCMC algorithm by combining Gibbs sampler and Metropolis-Hastings algorithm, which is mainly based on the ability of the B-splines to be expressed linearly and on the fact that the distribution of the model error can be obtained as scale mixture of normal distributions. We assume that a priori, the four parameters vectors (parametric and nonparametric components on location and dispersion submodels) are independent and normally distributed. The considered values for hyperparameters enable a direct comparison of the results with those obtained under the classical approach.

Value

`chains`	A matrix that contains the posterior sample of interest parameters. Each column represents the marginal posterior sample of each parameter.
`res`	a vector of quantile residuals, proposed by Dunn and Smyth (1996) in the context of classical inference, but suited here to the Bayesian case.
`K-L`	a vector of case-deletion influence measures based on the Kullback-Leibler divergence.
`X_2`	a vector of case-deletion influence measures based on the X2-Distance divergence.
`DIC`	DIC criterion for model selection.
`LMPL`	Log-marginal pseudo-likelihood for model selection.

Author(s)

Luz Marina Rondon <lumarp@gmail.com> and Heleno Bolfarine

References

Rondon, L.M. and Bolfarine, H. (2015) Bayesian Analysis of Generalized Elliptical Semi-parametric Models. (submitted).

Dunn, P.K. e Smyth, G.K. (1996). Randomized quantile residuals. Journal of Computational and Graphical Statistics. 5, 236-244.

See Also

bsp

Examples

```
##########   European Rabbit   ##########
#library(ssym)
#data(Erabbits)
#Erabbits2 <- Erabbits[order(Erabbits$age,Erabbits$wlens),]
#attach(Erabbits2)

#fit <- gesm(wlens ~ bsp(age) | bsp(age), family= "ContNormal", eta=c(0.8,0.9),
#  burn.in=1000, post.sam.s=5000, thin=10)
#summary(fit)

##########   Plot nonparametric components for the location and dispersion parameters
```

```
#par(mfrow=c(1,2))
#bsp.graph.gesm(fit, which=1, age, xlab="Rabbit age", ylab="f(age)", main="Location")
#bsp.graph.gesm(fit, which=2, age, xlab="Rabbit age", ylab="g(age)", main="Dispersion")

##########   Residual plot
#par(mfrow=c(1,2))
#plot(fit$res, ylim=c(-2,2), xlab="Index", ylab="", main="Residuals", cex=0.3,
#  type="p", lwd=3)
#abline(h=0,lty=3)
#qqnorm(fit$res, xlim=c(-2,2), ylim=c(-2,2), xlab="Quantile", ylab="Residuals",
#    cex=0.3, type="p", lwd=3)
#abline(0,1,lty=3)

##########   Influence measures plot
#par(mfrow=c(1,2))
#plot(fit$KL, xlab="Index", ylab="", main="K-L divergence", cex=0.3, type="p", lwd=3)
#abline(h=3*mean(fit$KL))
#plot(fit$X_2, xlab="Index", ylab="", main="X2 divergence", cex=0.3, type="p", lwd=3)
#abline(h=3*mean(fit$X_2))
```

---

mcmc.fmem *MCMC algorithm for Flexible Measurement Error Models*

Description

This function implements the MCMC algorithm derived in order to draw samples of the posterior distribution of the interest parameters in flexible measurement error models under the presence of homocedastic and heterocedastic random errors.

Usage

```
mcmc.fmem(params)
```

Arguments

params	An object type list, which provides the setup (i.e., values of hyperparameters, initial values, model matrices, basis functions for the B-splines, burn-in and posterior sample size) of the model requested by the user.

Author(s)

Luz Marina Rondon <lumarp@gmail.com> and Heleno Bolfarine

`mcmc.gesm` *MCMC algorithm for Generalized Elliptical Semi-parametric Models*

Description

This function implements the MCMC algorithm derived in order to draw samples of the posterior distribution of the interest parameters in generalized elliptical semi-parametric models, which combines Gibbs sampler and Metropolis-Hastings algorithm.

Usage

```
mcmc.gesm(params)
```

Arguments

`params`	An object type list, which provides the setup (i.e., values of hyperparameters, initial values, model matrices, basis functions for the B-splines, burn-in and posterior sample size) of the model requested by the user.

Author(s)

Luz Marina Rondon <lumarp@gmail.com> and Heleno Bolfarine

`summary.fmem` *Produces a complete summary of flexible measurement error model fit*

Description

summary.fmem displays the summary of the Bayesian analysis of flexible measurement error models. This function produces a table with the summary statistics (mean, median and standard deviation) of the posterior distribution as well as the 95% credible interval for the interest parameters. Further, this function displays goodness-of-fit statistics such as DIC and LMPL.

`summary.gesm` *Produces a complete summary of Generalized elliptical semi-parametric model fit*

Description

summary.gesm displays the summary of the Bayesian analysis of Generalized Elliptical Semi-parametric Models. This function produces a table with the summary statistics (mean, median and standard deviation) of the posterior distribution as well as the 95% credible interval for the interest parameters. Further, this function displays goodness-of-fit statistics such as DIC and LMPL.

`TexasData` *Relationship between income and demographic composition in Texas*

Description

This dataset consists of some demographic variables for each of the 254 municipalities of the state of Texas, USA, which were obtained from the US Census of 2010. From the *American Community Survey* (ACS), collected during the period 2009-2013, some variables of labor force and labor market of the households in each municipality of the state of Texas, were also included. Because the variables obtained from the ACS are subjected to sampling error, their analysis is based on the assumption of measurement error, where the sampling variances are considered as the variances of the measurement errors.

Usage

```
data(TexasData)
```

Format

A data frame with 254 observations on the following 13 variables.

`County` name of municipality

`MeanIng` Average family income in the last year, in hundreds of dollars

`VarMeanIng` Variance of average family income

`PopLabor16` Percentage of people in the Labor force (over 16 years).

`VarPopLabor16` Variance of percentage of people in the Labor force

`P_Desemp` Unemployment rate

`VarP_Desemp` Variance of unemployment rate

`PbMasc` Male population

`PbFem` Female population

`P_PbHisp` Percentage of Hispanic population

`P_PbNegra` Percentage of black population

`DensPobla` Population density

`P_PbFem` Percentage of female population

Source

http://www.census.gov/, http://www.census.gov/acs/www/

Examples

```
#data(TexasData)
#nu <- 3
#zeta <- nu/(nu-1)
#heter <- list(sigma2y=VarMeanIng/zeta, sigma2xi=cbind(VarPopLabor16,
#          VarP_Desemp)/zeta)
#model <- fmem(MeanIng/100 ~  PopLabor16 + P_Desemp | log(DensPobla) +
#         P_PbFem +  P_PNegra + bsp(P_PbHisp), data=TexasData, family="Slash",
#         eta=nu, burn.in=10000, post.sam.s=10000, heter=heter,thin=10)
#summary(model)
```

Index