Package 'BayesGESM'

January 27, 2015

Type Package

Title Bayesian Analysis of Generalized Elliptical Semiparametric Models and Flexible Measurement Error Models.

Version 1.2

Date 2015-01-09

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

Maintainer Luz Marina Rondon <lumarp@gmail.com>

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 semiparametric models, where both location and dispersion parameters of the response variable distribution include nonparametric additive components described by using B-splines; and (ii) Flexible measurement error models, which admit explanatory variables with and without measurement additive errors as well as the presence of a nonparametric component approximated by using B-splines.

License GPL-2 | GPL-3

Depends splines, truncdist, GIGrvg, normalp, Formula

Suggests ssym

NeedsCompilation no

Repository CRAN

Date/Publication 2015-01-09 06:06:24

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 nonparametric additive components described by using B-splines; and *(ii)* Flexible measurement error models, which admit explanatory variables with and without measurement additive errors as well as the presence of a nonparametric component approximated by using B-splines.

Details

Package:	BayesGESM
Type:	Package
Version:	1.2
Date:	2015-01-09
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= "Hyperbolic", eta=0.8,
#  burn.in=100, post.sam.s=100, thin=2)
#summary(fit)

###### Example for Flexible Measurement Error Models ################
#library(SemiPar)
#library(Formula)
#### Ragweed Pollen ####
#data(ragweed)
#attach(ragweed)
#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)
#
```

---

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 and has an intercept.

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 *Tool for plotting B-splines*

Description

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

Usage

```
bsp.graph(object, which, 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.
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 flexible measurement error models. This model admits vectors of explanatory variables with and without measurement error as well as the presence of a nonlinear effect, 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)
```

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`	the ratio $\omega = \sigma_y^2/\sigma_\xi^2$.
`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`	a numeric value or numeric vector that represents the extra parameter of the specified error distribution.
`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 of three parts, namely: *(i)* observed response variable; *(ii)* covariates with measurement error; and *(iii)* covariates without measurement error including the nonparametric component, 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 which admits vectors of explanatory variables with and without measurement error as well as the presence of a nonlinear effect 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

Examples

```
#library(SemiPar)
#### Ragweed Pollen ####
#data(ragweed)
#attach(ragweed)

#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 nonparametric component
#bsp.graph(model, which=1, xlab="Day", ylab="f(Day)")
#
```

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`	a numeric value or numeric vector that represents the extra parameter of the specified error distribution.
`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 component; and *(iii)* covariates for the dispersion parameter including the nonparametric component. 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.
EAIC	EAIC criterion for model selection.
EBIC	EBIC 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= "Hyperbolic", eta=0.8,
#  burn.in=100, post.sam.s=100, thin=2)
#summary(fit)

##########   Plot nonparametric components for the location and dispersion parameters
#par(mfrow=c(1,2))
#bsp.graph(fit,which=1, xlab="Rabbit age", ylab="f(age)", main="Location")
```

```
#bsp.graph(fit,which=2, 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.

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 Semiparametric 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 semiparametric 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 semiparametric model fit*

Description

summary.gesm displays the summary of the Bayesian analysis of Generalized Elliptical Semiparametric 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, EAIC, EBIC and LMPL.

Index