

Package ‘gllvm’

July 4, 2017

Type Package

Title Generalized Linear Latent Variable Models

Version 0.1.0

Author Jenni Niku, Francis K.C. Hui, Sara Taskinen, David I. Warton

Maintainer Jenni Niku <jenni.m.e.niku@jyu.fi>

Description Generalized linear latent variable model (GLLVM) for analyzing multivariate data.
Estimation is performed using either Laplace approximation or Variational approximation method.

License GPL-2

Imports mvtnorm, mvabund, MASS, Matrix, statmod, fishMod

Depends mvabund

Encoding UTF-8

LazyData true

RoxygenNote 6.0.1

URL <https://github.com/JenniNiku/gllvm.git>

R topics documented:

coef.gllvm	2
coefplot.gllvm	2
confint.gllvm	3
get.residual.cor.gllvm	4
gllvm	5
logLik.gllvm	9
ordplot.gllvm	10
print.gllvm	11
residuals.gllvm	11
sim_bacteria	13
summary.gllvm	13
Index	15

coef.gllvm	<i>Extract gllvm coefficients</i>
------------	-----------------------------------

Description

Extracts model coefficients from 'gllvm' objects.

Usage

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

Arguments

object	An object of class 'gllvm'
...	Not used.

Author(s)

David Warton, Jenni Niku <jenni.m.e.niku@jyu.fi>

Examples

```
## Not run:
library(mvabund) ## Load a dataset from the mvabund package
data(antTraits)
y <- as.matrix(antTraits$abund)
# Fit GLLVM model
fit <- gllvm(y = y, family = "negative.binomial")
# Coefficients
coef(fit)

## End(Not run)
```

coefplot.gllvm	<i>Plot covariate coefficients and confidence intervals</i>
----------------	---

Description

Plot covariate coefficients and their confidence intervals

Usage

```
coefplot.gllvm(fit, y.label = TRUE, which.Xcoef = NULL, cex.ylab = 1,
  mfrow = NULL, mar = c(4, 5, 2, 1), ...)
```

Arguments

<code>fit</code>	An object of class 'gllvm'
<code>y.label</code>	Logical, if TRUE (default) colnames of y with respect to coefficients are added to plot.
<code>which.Xcoef</code>	vector indicating which X-coefficients will be plotted. Can be vector of covariate names or numbers. Default is NULL when all covariate coefficients are plotted.
<code>cex.ylab</code>	The magnification to be used for axis annotation relative to the current setting of cex.
<code>mfrow</code>	Same as mfrow in par. If NULL (default) it is determined automatically.
<code>mar</code>	vector of length 4, which defines the margin sizes: c(bottom, left, top, right). Defaults to c(4, 5, 2, 1).
<code>...</code>	Additional graphical arguments.

Details

Plot covariate coefficients and their confidence intervals

Author(s)

Jenni Niku <jenni.m.e.niku@jyu.fi>, Francis K.C. Hui, Sara Taskinen

Examples

```
## Not run:
library(mvabund) ## Load a dataset from the mvabund package
data("antTraits")
y <- antTraits$abund
X <- antTraits$env
TR <- antTraits$traits
X <- scale(X)

fitx <- gllvm(y = y, X = X, family = "negative.binomial")
coefplot.gllvm(fitx, mfrow=c(3,2), mar = c(4,8,2,1), cex.ylab=0.8)
coefplot.gllvm(fitx, which.Xcoef = c("Bare.ground", "Shrub.cover"),
               mar = c(4,8,2,1), cex.ylab = 0.8)

fitT <- gllvm(y = y, X = X, TR = TR, family = "negative.binomial")
coefplot.gllvm(fitT, mar = c(4,8,2,1))

## End(Not run)
```

confint.gllvm

Confidence intervals for model parameters

Description

Computes confidence intervals for parameters in a fitted GLLVM model.

Usage

```
## S3 method for class 'gllvm'
confint(object, level = 0.95, ...)
```

Arguments

object	An object of class 'gllvm'.
level	The confidence level. Scalar between 0 and 1.
...	Not used.

Author(s)

Jenni Niku <jenni.m.e.niku@jyu.fi>

Examples

```
## Not run:
library(mvabund) ## Load a dataset from the mvabund package
data(antTraits)
y <- as.matrix(antTraits$abund)
X <- as.matrix(antTraits$env[,4:5])
# Fit GLLVM model
fit <- gllvm(y = y, X = X, family = "negative.binomial")
# 95 % confidence intervals
confint(fit, level = 0.95)

## End(Not run)
```

```
get.residual.cor.gllvm
```

Extract residual correlations from gllvm object

Description

Calculates the residual correlation from 'gllvm' models that include latent variables.

Usage

```
get.residual.cor.gllvm(object)
```

Arguments

object	An object of class 'gllvm'
--------	----------------------------

Author(s)

David Warton

Examples

```
## Not run:
library(mvabund) ## Load a dataset from the mvabund package
data(antTraits)
y <- as.matrix(antTraits$abund)
# Fit GLLVM model
fit <- gllvm(y = y, family = "negative.binomial")
# residual correlations:
```

```
get.residual.cor.gllvm(fit)

## End(Not run)
```

gllvm

Generalized Linear Latent Variable Models.

Description

Fit generalized linear latent variable models for multivariate data. The models can be fitted using Laplace approximation method or Variational approximation method.

Usage

```
gllvm(y, X = NULL, TR = NULL, data = NULL, n.X = 0, n.TR = 0,
      num.lv = 2, family, method = "VA", row.eff = FALSE, get.fourth = TRUE,
      get.trait = TRUE, offset = NULL, start.lvs = NULL,
      start.params = NULL, sd.errors = TRUE, Lambda.struc = "unstructured",
      diag.iter = 5, trace = FALSE, plot = FALSE, la.phi.upd = "phi",
      la.link.bin = "logit", n.init = 1, fixed.power = TRUE, Power = 1.5,
      constrOpt = FALSE, restrict = 30, eps = 1e-05, seed = NULL,
      max.iter = 200, maxit = 1000)
```

Arguments

y	(n x m) matrix of responses. If data is in long format, y is a data.frame including a row index vector, a column index vector and a vector of response variables in this order.
X	Matrix of environmental covariates.
TR	Matrix of trait covariates. Can be used only with VA method
data	Matrix of environmental and trait covariates. Can be used only if responses y are in long format.
n.X	Number of environmental covariates in data matrix above.
n.TR	Number of trait covariates in data matrix above.
num.lv	Number of latent variables, d in GLLVM model. Non-negative integer. Defaults to 2.
family	Distribution function to be used in the model. Options are "poisson" (with log link), "negative.binomial" (with log link) and "ZIP" (with log link), "binomial" (with logit/cloglog link when method="LA" and probit link when method="VA"), Tweedie (with log link, only with "LA"-method), "ordinal" (only with "VA"-method).
method	Model can be fitted using Laplace Approximation method (method="LA") or Variational Approximation method (method="VA"). Defaults to "VA".
row.eff	Logical, indicating whether row effects are included in the model. Defaults to FALSE.
get.fourth	Logical. If TRUE (default) fourth corner/interaction terms of environmental and trait covariates are included in the model.

get.trait	Logical. If TRUE (default) main effects of trait covariates are included in the model.
offset	Vector or matrix of offset terms
start.lvs	Initialize starting values for latent variables with (n x num.lv) matrix. Defaults to NULL when starting values are generated randomly.
start.params	object of class 'gllvm' which can be given as starting parameters for count data (poisson, NB, or ZIP).
sd.errors	Logical. If TRUE (default) standard errors for parameter estimates are calculated.
Lambda.struc	Covariance structure for latent variables when method = "VA", "unstructured" or "diagonal".
diag.iter	Non-negative integer which is used to speed up the updating of variational parameters Lambda in VA method. Defaults to 5.
trace	Logical, if TRUE in each iteration step information on current step will be printed. Defaults to FALSE.
plot	Logical, if TRUE ordination plots will be printed on each iteration step. Defaults to FALSE.
la.phi.upd	Defaults to "phi" when dispersion parameters with LA method are updated using parametrization ϕ . If "inv.phi", dispersion parameters are updated using parametrization $1/\phi$.
la.link.bin	Link function for binomial family if method="LA". Options are "logit" and "cloglog".
n.init	Number of initial runs. Uses multiple runs and picks the one giving highest log-likelihood value. This is recommendable because the method is quite sensitive to initial values of the latent variables. Defaults to 1.
fixed.power	Logical, if TRUE (default) the power parameter in Tweedie model is fixed.
Power	Fixed power parameter if fixed.power=TRUE or starting value for power parameter if fixed.power=FALSE in Tweedie model. Scalar from interval (1,2). Defaults to 1.5.
constrOpt	Logical, if TRUE parameters are estimated by constraining linear predictors in optimization. This is recommendable if algorithm is very unstable in different runs. Defaults to FALSE. Cannot be used when family is "tweedie" or "ZIP". Constraint for the absolute value of the linear predictor is defined by restrict.
eps	Convergence criteria for log likelihood, defaults to 1e-4.
seed	a single seed value, defaults to NULL.
max.iter	Maximum number of iterations, defaults to 100.
maxit	Maximum number of iterations within optim function, defaults to 1000.

Details

Fits generalized linear latent variable models. Method can be used with two types of latent variable models depending on covariates. If only site related environmental covariates are used, the expectation of response Y_{ij} is determined by

$$g(\mu_{ij}) = \eta_{ij} = \alpha_i + \beta_{0j} + x'_i \beta_j + u'_i \theta_j,$$

where $g(\cdot)$ is a known link function, u_i are d -variate latent variables, α_i is an optional row effect at site i , β_{0j} is intercept term for species j , β_j and θ_j are column specific coefficients related to

covariates and the latent variables, respectively. If also trait covariates are included, the expectation of response Y_{ij} is

$$g(\mu_{ij}) = \alpha_i + \beta_{0j} + x'_i\beta_x + TR'_j\beta_t + \text{vec}(B) * \text{kronecker}(TR_j, X_i) + u'_i\theta_j$$

where $g(\cdot)$, u_i , β_{0j} and θ_j are defined as above. Vectors β_{x_t} and β_{t_t} are main effects or coefficients related to environmental and trait covariates, respectively, matrix B includes interaction terms. This is the so called 'fourth corner model', but now interaction/fourth corner terms are optional as well as are the main effects of trait covariates. The fourth corner model is implemented only using variational approximation method.

Distributions: Mean and variance for distributions are defined as follows.

For count data family="poisson": Expectation $E[Y_{ij}] = \mu_{ij}$, variance $V(\mu_{ij}) = \mu_{ij}$. OR family="negative.binomial": Expectation $E[Y_{ij}] = \mu_{ij}$, variance $V(\mu_{ij}) = \mu_{ij} + \phi_j * \mu_{ij}^2$ OR family="ZIP": Expectation $E[Y_{ij}] = (1 - p)\mu_{ij}$, variance $V(\mu_{ij}) = \mu_{ij}(1 - p)(1 + \mu_{ij}p)$

For binary data family="binomial": Expectation $E[Y_{ij}] = \mu_{ij}$, variance $V(\mu_{ij}) = \mu_{ij}(1 - \mu_{ij})$.

For biomass data family="tweedie": Expectation $E[Y_{ij}] = \mu_{ij}$, variance $V(\mu_{ij}) = \phi_j * \mu_{ij}^\nu$, where ν is a power parameter of Tweedie distribution. See details Dunn and Smyth (2005).

For ordinal data family="ordinal": See Hui et.al. (2016).

- Method is sensitive for the choices of initial values of the latent variables. Therefore it is recommendable to use multiple runs and pick up the one giving the highest log-likelihood value. However, if this is computationally too demanding you can use the result of some classical ordination method (for example nMDS) as initial values for latent variables.

- If algorithm is very unstable in different runs, it is recommendable to use constrOpt=TRUE. The parameters are then estimated by constraining linear predictors in optimization. This has been done using function constrOptim. Especially with binomial family, algorithm may often converge to a poor local maximum and constrained optimization may help in such cases.

Value

An object of class "gllvm" includes the following components:

call	function call
logL	log likelihood
lvs	latent variables
params	list of parameters <ul style="list-style-type: none"> • theta coefficients related to latent variables • \$beta0 column specific intercepts • \$Xcoef coefficients related to environmental covariates X • \$Tcoef coefficients related to trait covariates TR • \$fourth interaction terms • \$row.params row specific intercepts • \$phi dispersion parameters ϕ for negative binomial or Tweedie family • \$inv.phi dispersion parameters $1/\phi$ for negative binomial • \$p Probability of zero inflation for ZIP family
Power	power parameter ν for Tweedie family
sd	list of standard errors of parameters

Author(s)

Jenni Niku <jenni.m.e.niku@jyu.fi>, Francis K.C. Hui, Sara Taskinen

References

- Dunn, P. K. and Smyth, G. K. (2005). Series evaluation of tweedie exponential dispersion model densities. *Statistics and Computing*, 15:267-280.
- Hui, F. K. C., Taskinen, S., Pledger, S., Foster, S. D., and Warton, D. I. (2015). Model-Based Approaches to Unconstrained Ordination. *Methods in Ecology and Evolution*, 6:399-411.
- Hui, F. K. C., Warton, D., Ormerod, J., Haapaniemi, V., and Taskinen, S. (2016). Variational Approximations for Generalized Linear Latent Variable Models. *Journal of Computational and Graphical Statistics*, 26:35-43.
- Niku, J., Warton, D. I., Hui, F. K. C., and Taskinen, S. (2017). Generalized Linear Latent Variable Models for Multivariate Count and Biomass Data in Ecology. Under revision.
- Warton, D. I., Guillaume Blanchet, F., O'Hara, R. B., Ovaskainen, O., Taskinen, s., Walker, S. C. and Hui, F. K. C. (2015). So Many Variables: Joint Modeling in Community Ecology. *Trends in Ecology & Evolution*, 30:766-779.

See Also

[ordplot.gllvm](#), [summary.gllvm](#), [residuals.gllvm](#), [confint.gllvm](#), [coefplot.gllvm](#).

Examples

```
## Not run:
library(mvabund) ## Load a dataset from the mvabund package
data(antTraits)
y <- as.matrix(antTraits$abund)
X <- as.matrix(antTraits$env[,4:5])
TR <- antTraits$traits[,4:5]

## Example 1: Fit model with two latent variables
# Using Laplace approximation:
fitl0 <- gllvm(y=y,family = "negative.binomial", method="LA",seed=1)
ordplot.gllvm(fitl0)
# Using Variational analysis:
fitv0 <- gllvm(y=y,family = "negative.binomial", method="VA",seed=1)
ordplot.gllvm(fitv0)

# Poisson family:
fit.p <- gllvm(y=y,family = "poisson", method="LA")
ordplot.gllvm(fit.p)
# Use poisson model as a starting parameters for ZIP-model:
fit.z <- gllvm(y=y,family = "ZIP", method="LA",start.param = fit.p)
ordplot.gllvm(fit.z)

## Example 2: Fit model with two latent variables and environmental covariates,
fitlX <- gllvm(y=y, X = X,family = "negative.binomial", method="LA")
ordplot.gllvm(fitlX)
# Use 5 initial runs and pick the best one
fitlX_5 <- gllvm(y=y, X = X,family = "negative.binomial", method="LA")
ordplot.gllvm(fitlX_5)
```



```
# Example 3: Fit fourth corner model with two latent variables
# This can be fitted using only Variational analysis:
fitF <- gllvm(y=y, X = X, TR=TR,family = "negative.binomial", method="VA")
ordplot.gllvm(fitF)

# Example 4: Fit Tweedie model for coral data
data(tikus)
ycoral <- tikus$abund
# Exclude species which have observed at less than 5 sites
ycoral <- ycoral[(colSums(ycoral>0)>4)]
fit.twe <- gllvm(y=ycoral,family = "tweedie", method="LA")
ordplot.gllvm(fit.twe)

## End(Not run)
```

logLik.gllvm

*Log-likelihood of gllvm***Description**

Extracts Log-likelihood from 'gllvm' objects.

Usage

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

Arguments

object	An object of class 'gllvm'
...	Not used.

Author(s)

David Warton

Examples

```
## Not run:
library(mvabund) ## Load a dataset from the mvabund package
data(antTraits)
y <- as.matrix(antTraits$abund)
# Fit GLLVM model
fit <- gllvm(y = y, family = "negative.binomial")
# log-Likelihood:
logLik(fit)

## End(Not run)
```

ordplot.gllvm	<i>Plot latent variables from GLLVM model.</i>
---------------	--

Description

Plot latent variables and their corresponding coefficients (biplot).

Usage

```
ordplot.gllvm(object, biplot = FALSE, ind.spp = NULL, alpha = 0.5,
  main = NULL, which.lvs = c(1, 2), jitter = FALSE, ...)
```

Arguments

object	An object of class 'gllvm'
biplot	TRUE if both latent variables and their coefficients are plotted, FALSE if only LVs
ind.spp	Number of LV coefficients included in the biplot.
alpha	A numeric scalar between 0 and 1 that is used to control the relative scaling of the latent variables and their coefficients, when constructing a biplot.
which.lvs	Indices of two latent variables to be plotted if number of the latent variables is more than 2. A vector with length of two. Defaults to c(1, 2).
jitter	if TRUE, jittering is applied on points
...	Additional graphical arguments.

Details

Construct a scatter plot of two latent variables, i.e. an ordination plot. If only one latent variable is in the fitted model, latent variables are plotted against their corresponding row indices. The latent variables are labeled using the row index of the response matrix *y*.

Coefficients related to latent variables are plotted in the same figure with the latent variables if 'biplot=TRUE'. They are labeled using the column names of *y*. The number of latent variable coefficients to be plotted can be controlled by *ind.spp*. An argument *alpha* is used to control the relative scaling of the latent variables and their coefficients. If *alpha*=0.5, the latent variables and their coefficients are on the same scale.

Author(s)

Jenni Niku <jenni.m.e.niku@jyu.fi>, Francis K.C. Hui

Examples

```
## Not run:
library(mvabund) ## Load a dataset from the mvabund package
data(antTraits)
y <- as.matrix(antTraits$abund)
# Fit GLLVM model
fit <- gllvm(y = y, family = "negative.binomial")
# Ordination plot:
ordplot.gllvm(fit)
# Biplot with 10 species
```

```
ordplot.gllvm(fit, biplot = TRUE, ind.spp = 10)

## End(Not run)
```

print.gllvm	<i>Print GLLVM</i>
-------------	--------------------

Description

Print an object of class 'gllvm'

Usage

```
## S3 method for class 'gllvm'
print(x, ...)
```

Arguments

x	An object of class 'gllvm'
...	Not used.

Author(s)

Jenni Niku <jenni.m.e.niku@jyu.fi>

Examples

```
## Not run:
library(mvabund) ## Load a dataset from the mvabund package
data(antTraits)
y <- as.matrix(antTraits$abund)
# Fit GLLVM model
fit <- gllvm(y = y, family = "negative.binomial")
# Print:
print(fit)

## End(Not run)
```

residuals.gllvm	<i>Dunn-Smyth -residuals for GLLVM model</i>
-----------------	--

Description

Calculates Dunn-Smyth -residuals for GLLVM model and plots them.

Usage

```
## S3 method for class 'gllvm'
residuals(object, plot = TRUE, by.index = NULL,
  xlim = NULL, ...)
```

Arguments

object	An object of class 'gllvm'
plot	Logical, construct a plot of residuals. Defaults to FALSE
by.index	If NULL (default), residuals are plotted by linear predictors. If by.index="site", residuals are plotted by row index. If by.index="spp" residuals are plotted by column index.
...	Additional graphical arguments.

Details

Computes Dunn-Smyth residuals or randomized quantile residuals (Dunn and Smyth, 1996) for GLLVM model.

Value

A list containing residuals which is a matrix of residuals and linpred which is a matrix of linear predictors.

Author(s)

Jenni Niku <jenni.m.e.niku@jyu.fi>

References

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

Examples

```
## Not run:
library(mvabund) ## Load a dataset from the mvabund package
data(antTraits)
y <- as.matrix(antTraits$abund)
# Fit GLLVM model
fit <- gllvm(y = y, family = "negative.binomial")
# residuals saved and plotted by linear predictors
res <- residuals(fit, plot = TRUE, by.index = NULL)
# residuals plotted by site index
res <- residuals(fit, plot = TRUE, by.index = "site")
# residuals plotted by column index
res <- residuals(fit, plot = TRUE, by.index = "spp")

## End(Not run)
```

sim_bacteria	<i>Simulated bacteria data</i>
--------------	--------------------------------

Description

Simulated dataset of size 56 x 985 that can be used to illustrate the methods in Section 5.1 of Niku et al., (2017).

Usage

```
data(sim_bacteria)
```

Arguments

y	A data frame with abundances of 985 species measured at 56 sites
X	A matrix of six environmental variables at each of the 56 sites

References

Niku, J., Warton, D. I., Hui, F. K. C., and Taskinen, S. (2017). Generalized Linear Latent Variable Models for Multivariate Count and Biomass Data in Ecology. Under revision.

summary.gllvm	<i>Summary of fitted gllvm object</i>
---------------	---------------------------------------

Description

A summary of the fitted gllvm object, including function call, distribution family and model parameters.

Usage

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

Arguments

object	An object of class 'gllvm'
...	Not used.

Author(s)

Jenni Niku <jenni.m.e.niku@jyu.fi>

Examples

```
## Not run:
library(mvabund) ## Load a dataset from the mvabund package
data(antTraits)
y <- as.matrix(antTraits$abund)
# Fit GLLVM model
fit <- gllvm(y = y, family = "negative.binomial")
summary(fit)

## End(Not run)
```

Index

`coef.gllvm`, [2](#)
`coefplot.gllvm`, [2](#), [8](#)
`confint.gllvm`, [3](#), [8](#)

`get.residual.cor.gllvm`, [4](#)
`gllvm`, [5](#)

`logLik.gllvm`, [9](#)

`ordplot.gllvm`, [8](#), [10](#)

`print.gllvm`, [11](#)

`residuals.gllvm`, [8](#), [11](#)

`sim_bacteria`, [13](#)
`summary.gllvm`, [8](#), [13](#)