Package ‘RCPmod’

February 19, 2015

Version 1.4
Date 2013-11-01
Title Regions of common profiles modelling with mixtures-of-experts.
Author Scott D. Foster
Maintainer Scott Foster <scott.foster@csiro.au>
Depends numDeriv, gtools
Suggests MASS
Description Identifies regions of common (species) profiles (RCPs). Within a region the probability of sampling all species remains approximately constant. This is performed using mixtures-of-experts models. The package also contains associated methods, such as diagnostics. Currently, the model and methods are defined for species presence-absence data for a number of sites.
License GPL (>= 2)
NeedsCompilation yes
Repository CRAN
Date/Publication 2013-11-02 10:53:28

R topics documented:

AIC.regimix .......................................................... 2
BIC.regimix .......................................................... 2
cof.regimix ............................................................ 3
extractAIC.regimix .................................................. 4
logLik.regimix ......................................................... 4
orderFitted ........................................................... 5
plot.regimix ........................................................... 5
predict.regimix ........................................................ 6
print.regimix .......................................................... 7
regimix ................................................................. 8
simRCPdata .......................................................... 10
summary.regimix ...................................................... 12
vcov.regimix .......................................................... 12

Index 14
Description

Returns information criterion for regimix models.

Arguments

object an object obtained from fitting a region of common profile mixture model. Such as that generated from a call to regimix(qv).

k the coefficient for the penalty in the information criterion. k=2 signifies Akaike’s information criterion, k=log(object$n) corresponds to the Bayesian information criterion. If NULL (default) the AIC is used.

... ignored

Value

A numeric scalar giving the information criterion.

Method

AIC( object, ..., k=2)

Author(s)

Scott D. Foster

Description

Returns information criterion for regimix models.

Arguments

object an object obtained from fitting a region of common profile mixture model. Such as that generated from a call to regimix(qv).

... ignored

Value

A numeric scalar giving the Bayesian information criterion.
**Method**

BIC( object, ... )

**Author(s)**

Scott D. Foster

---

```r
coef.regimix

A regimix objects coefficients.
```

**Description**

Returns coefficients from a regimix object.

**Arguments**

- `object`:
  - An object obtained from fitting a regions of common profile mixture model. Such as that generated from a call to `regimix(qv)`.  
- `...`:
  - Ignored

**Value**

Returns a list of three elements, one each for the estimates for the species prevalence (alpha), the deviations from alpha for the first (nRCP-1) regional profiles (tau), and the (nRCP-1) sets of region regression coefficients (beta).

**Method**

`coef( object, ... )`

**Author(s)**

Scott D. Foster
extractAIC.regimix  
Extracts the AIC for a fitted regimix model.

Description

Computes the generalised AIC for a regimix model.

Arguments

- **fit**: an object obtained from fitting a region of common profile mixture model. Such as that generated from a call to regimix(qv).
- **scale**: ignored
- **k**: the coefficient for the penalty in the information criterion. \( k=2 \) corresponds to Akaike's information criterion, \( k=\log(\text{object}$n) \) corresponds to the Bayesian information criterion. Default is \( k=2 \) (AIC).
- **...**: ignored

Value

A two element numeric vector. First element is the number of parameters in the model. The second is the information criterion.

Method

extractAIC( fit, scale, k=2, ... )

Author(s)

Scott D. Foster

logLik.regimix  
The log likelihood for a regimix model.

Description

Returns the maximised log likelihood for a regimix model.

Arguments

- **object**: an object obtained from fitting a regimat mixture model. Such as that generated from a call to regimix(qv).
- **...**: ignored
Value
A numeric scalar giving the maximised log likelihood for the regimix model.

Method
logLik( object, . . . )

Author(s)
Scott D. Foster

orderFitted
Assesses classification error and re-orders parameters

Description
Not to be used by general public – unless you know what the function actually does.

Author(s)
Scott D. Foster

plot.regimix
Plots a regimix object.

Description
Plots the residuals, along with simulation envelopes, from a regimix object.

Arguments

x 
an object obtained from fitting a RCP mixture model. Such as that generated from a call to regimix(qv).

nsim 
the number of simulations to use for the confidence interval. Default is 100, serious usage is likely to require more.

alpha.conf 
the confidence level(s) to use in the residual plots. Default is c(0.90,0.95,0.99)

Details
Produces a plot with two panels. The first is for point-wise confidence intervals and the second is for approximate global intervals. See Foster et al (2012) for details.
Method

\begin{verbatim}
plot( x, ..., nsim=100, alpha.conf=c(0.90,0.95,0.99))
\end{verbatim}

Author(s)

Scott D. Foster

References


\begin{verbatim}
predict.regimix Predicts from a regimix object.
\end{verbatim}

Description

Predicts RCP probabilities at a series of sites. Confidence intervals are available too.

Arguments

- **object**
  - an object obtained from fitting a RCP mixture model. Such as that generated from a call to regimix(qv).

- **newdata**
  - a data.frame (or something that can be coerced) containing the values of the covariates where predictions are to be made. If NULL (the default) then predictions are made at the locations of the original data.

- **nboot**
  - the number of parametric bootstrap samples to take for the bootstrap predictions, standard errors and confidence intervals. The default is 0, that is no bootstrapping is to be done and point predictions only are given.

- **alpha**
  - a numeric within [0,1] (well [0.5,1] really) indicating the specified confidence for the confidence interval. Argument is redundant if nboot == 0.

- **...**
  - ignored

Details

This function can take a while to run – it is a bootstrap function. nboot samples of the parameters are taken and then used to predict at each set of covariates defined in newdata. Quantiles of the resulting sets of bootstrap predictions are then taken. It is the last step that really takes a while.

Value

If nboot==0 then a n x H matrix of prior predictions (n=nrow(newdata), H=number of RCPs). Each row should sum to one.

if nboot!=0 then a list is returned. It has elements:
print.regimix

ptPreds the n x H matrix of point predictions
bootPreds the n x H matrix of bootstrap point predictions (mean of bootstrap samples)
bootSEs the n x H matrix of bootstrap standard errors for predictions
bootCIs the n x H x 2 array of bootstrap confidence intervals. Note that bootCIs[,1] gives the lower CIs and bootCIs[,2] gives the upper CIs.

Method

predict( object, ..., newdata=NULL, nboot=0, alpha=0.95)

Author(s)

Scott D. Foster

print.regimix

Prints a regimix object.

Description

Prints a regimix object. It may be useful for something.

Arguments

x an object obtained from fitting a RCP mixture model. Such as that generated from a call to regimix(qv).

Details

A list is returned that will be printed on exit, if not assigned to anything. It contains the function call and the estimated coefficients.

Method

print( x, ...)
regimix

Fits a regimix model.

Description

Fits a mixture-of-experts model to identify regions of similar community composition.

Usage

regimix( form=NULL, data, nRCP=3, control=list(), inits="random")

Arguments

form an object of class "formula" (or one that can be coerced to that class). The left hand side of this formula specifies the columns of the data argument that are the species binary data. The right hand side of this formula specifies the dependence of the region of common profile (RCP) probabilities on covariates. An example formula is cbind(spp1,spp2,spp3)~1+poly(temp,3) where spp1, spp2 and spp3 are species labels (columns of data), and the RCP probabilities depend on a (transformed) cubic polynomial of temp. No offset term can be accounted for, please do not enter one – the behaviour is unknown and could cause your computer to have awkward moments.

data an object of class "data.frame" (or one that can be coerced to that class). This data.frame contains all the data for the terms in the form argument

nRCP an integer (or something that can be coerced to an integer). This argument specifies the number of RCPs that will be fitted.

control a list of control parameters for optimisation and calculation. See details.

inits either a character string or a numeric vector. If a character string ("hclust" or "random") then it gives the method to generate starting values. If a numeric vector then it specifies the values of alpha, tau, and beta (in that order). It will be used unchecked if not a character.

Details

A typical formula for use in the form argument will have the form (for example) cbind(spp1,spp2,spp3,spp4)~1+cov1+cov2*cov3. This signifies that there are 4 species to be used for RCP modelling and that the RCP types are dependent on cov1+cov2+cov3+cov2: cov3. See ?glm for a description of how the right hand side of the formula is expanded.

The computation strategy for the default method, which has been demonstrated to work, is fully described in Foster et al (2012).

Control arguments for optimisation generally follow those in optim(qv), although a few differences occur (e.g. maxitInner). The elements of the control list are

maxit The maximum number of iterations. Default is 500.
maxGSiters  The maximum number of Gauss-Seidel iterations in a hybrid Gauss-Seidel – quasi-Newton optimisation scheme. This specifies the number of Gauss-Seidel iterations to use before switching to quasi-Newton. Default is 0.

maxitInner  The maximum number of iterations for each update within a Gauss-Seidel update. Default is 100. This argument is redundant if maxGSiters=0.

trace  Non-negative integer. If positive, tracing information on the progress of the optimization is produced. Higher values may produce more tracing information.

nreport  The frequency of reports for optimisation. Default is 1 – a report for every iteration.

reitol  Relative convergence tolerance. The algorithm stops if it is unable to reduce the value by a factor of reitol * (abs(val) + reitol) at a step. Defaults to sqrt(.Machine$double.eps), typically about 1e-8.

loglOnly  Should only the log-likelihood be cauculated? If TRUE then optimisation does not occur. If FALSE (default) then optimisation is performed.

penalty  A numeric scalar. This is the concentration for the Dirichlet-inspried penalty for the prior probabilities. Values less than zero will be set to the default, 0.01.

Value

regimix returns an object of class regimix. The regimix class has several methods: coef, plot, predict, residuals, summary, and vcov. The regimix object consists of a list with the following elements:

n  the number of samples.
S  the number of species.
p  the number of covariates used to delineate the regitats.
nRCP  the number of RCPs
logl  the maximised log likelihood.
coefs  a list of three elements, one each for the estimates for the species prevalence (alpha), the deviations from alpha for the first (nRCP-1) RCP (tau), and the (nRCP-1) sets of RCP regression coefficents (beta).
scores  a list of three elements. Structure corresponds to coefs.
pis  an n x nRCP matrix with each column giving the prior probabilities for the corresponding RCP type. Rows sum to one.
postProbs  an n x nRCP matrix with each column giving the posterior probabilities for the corresponding RCP type. Rows sum to one.
mus  a matrix of size nRCP x S where each row is the probability of each species being observed in that RCP type.
residuals  deviance residuals for each site.
BIC  Bayesian information criterion for the maximised model.
AIC  Akaike an information criterion for the maximised model.
start.vals  the values used to start the estimation procedure.
call  the call to the function.
titbits  other pieces of information, useful to developers, that users should not need to concern themselves with.
simRCPdata

Author(s)
Scott D. Foster

References

Examples
#simulate data
eexample( simRCPdata)
#fit the model
my.form <- paste( paste( paste(
   'cbind(', paste( paste( 'spp', 1:S, sep=''), collapse=''), sep=''),
   '),sep=''),
   '-x1.1+x1.2+x1.3+x2.1+x2.2+x2.3',sep=''))
my.cont <- list( maxGSsites = 0, maxit=1000, nreport=10, penalty=0.01)
fm <- regimix(form = my.form, data = simDat, nRCP = 3, control = my.cont, inits = "hclust")

simRCPdata Simulates from a regimix model

Description
Simulates a data set from a mixture-of-experts model for RCP (for region of common profile) types.

Usage
simRCPdata(H=3, S=20, p=3, n=200, alpha=NULL, tau=NULL, beta=NULL, X=NULL)

Arguments

H Integer giving the number of RCPs
S Integer giving the number of species
p Integer giving the number of covariates (including the intercept)
n Integer giving the number of observations (sites)
alpha Numeric vector of length S. Specifies the mean prevalence for each species, on the logit scale.
tau Numeric matrix of dimension c(H-1,S). Specifies each species difference from the mean to each RCPs mean for the first H-1 RCPs. The last RCP means are calculated using the sum-to-zero constraints
beta Numeric matrix of dimension c(H-1,p). Specifies the RCP’s dependence on the covariates (in X)
X Numeric matrix of dimension c(n,p). Specifies the covariates for the RCP model. Must include the intercept, if one is wanted
**Value**

A data frame that contains the outcomes (species data) and the covariates (environmental data). This data.frame has a number of special attributes, which are information about the model underlying the data. They are:

- **rcps**  
  the true, but unobserved, RCP types

- **pis**  
  the true prior probabilities

- **alpha**  
  the species overall prevalences, on linear predictor scale

- **tau**  
  the deviation from alpha for each RCP type, on linear predictor scale

- **beta**  
  the parameters controlling how the RCP types depend on the covariates

- **mu**  
  the probabilities of each species occurring in each RCP type

**Author(s)**

Scott D. Foster

**References**


**Examples**

```r
#generates synthetic data in Foster et al (2012)
simRCPdata

set.seed(151)
n <- 150
S <- 100
H <- 3
V <- as.data.frame( cbind( x1=runif( n, min=-10, max=10), x2=runif( n, min=-10, max=10) ))
pols <- list()
pols[[1]] <- poly( X$x1, degree=3)
pols[[2]] <- poly( X$x2, degree=3)
V <- as.matrix( cbind( 1, X, pols[[1]], pols[[2]]))

colnames(V) <- c("const", 'x1', 'x2', paste("x1",1:3,sep='.'), paste("x2",1:3,sep='.'))
p <- ncol( V[,2:3])
alpha <- rnorm( S)
tau.var <- 0.5
b <- sqrt( tau.var/2)
tau <- matrix( rexp( n=(H-1)*S, rate=1/b) - rexp( n=(H-1)*S, rate=1/b), nrow=H-1, ncol=S)
beta <- 0.2 * matrix( c(-1.2, -2.6, 0.2, -23.4, -16.7, -18.7, -59.2, -76.0, -14.2, -28.3, 
  -36.8, -17.8, -92.9,-2.7), nrow=H-1, ncol=p)
set.seed(121)
simDat <- simRCPdata( H=H, S=S, p=p, n=n, alpha=alpha, tau=tau, beta=beta, X=V[,2:3])
```
summary.regimix  
**Summarises a regimix object.**

**Description**

A summary from a regimix object. It may be useful for something.

**Arguments**

- **object**  
a object obtained from fitting a RCP (for region of common profile) mixture model. Such as that generated from a call to regimix(qv).
- **...**  
ignored

**Details**

A table is printed that contains the coefficient values, their standard errors, and their z-statistic. The second and third columns may be unreliable for some parameters.

**Method**

summary( object, ...) 

**Author(s)**

Scott D. Foster

vcov.regimix  
**Variance matrix for a regimix object.**

**Description**

Calculates variance-covariance matrix from a regimix object.

**Arguments**

- **object**  
an object obtained from fitting a RCP (for region of common profile) mixture model. Such as that generated from a call to regimix(qv).
- **method**  
the method to calculate the variance-covariance matrix. Options are: `simple` (default) and Richardson from the numDeriv package (finite difference approximations, see jacobian(qv) from numDeriv package); EmpInfo; and boot for empirical bootstrapping. The EmpInfo and boot methods are purely experimental. In fact, they are unlikely to work. The simple and Richardson methods will work for the regression parameters and may work for the species specific parameters.
- **nboot**  
the number of bootstrap samples to take for the bootstrap estimation. Argument is ignored if method != boot
- **...**  
ignored
Details

If method is either `simple` or `Richardson`, then the estimates variance matrix is based on a finite difference approximation to the observed information matrix. This method appears to be the most robust for the parameters relating the RCP types to the covariates.

If method is `EmpInfo` then an empirical approximation to the Fisher Information is used as the basis for the variance matrix estimate. These variances are only available for the parameters that relate the RCP types to the covariates. This method is experimental! Initial results are not good.

If method is `boot` then an empirical bootstrap is used to estimate the variance matrix. This method is experimental and initial results suggest that the resulting variances are downwardly biased.

Value

A square matrix of size equal to the number of parameters. It contains the variance matrix of the parameter estimates. If method=='EmpInfo', then only the sub-matrix corresponding to those parameters relating the RCP types to the covariates have non-NA elements.

Method

vcov( object, ..., method='simple', nboot=100)

Author(s)

Scott D. Foster
Index

*Topic **misc**

- AIC.regimix, 2
- BIC.regimix, 2
- coef.regimix, 3
- extractAIC.regimix, 4
- logLik.regimix, 4
- orderFitted, 5
- plot.regimix, 5
- predict.regimix, 6
- print.regimix, 7
- regimix, 8
- simRCPdata, 10
- summary.regimix, 12
- vcov.regimix, 12