Package 'CoDaLoMic'

June 13, 2025

Type Package

Title Compositional Models to Longitudinal Microbiome Data

Version 0.1.1

Author Irene Creus Martí [aut, cre] (ORCID: <`https://orcid.org/0000-0002-7962-4478`>)

Maintainer Irene Creus Martí <`ircrmar@mat.upv.es`>

Description Implementation of models to analyse compositional microbiome time series taking into account the interaction between groups of bacteria. The models implemented are described in Creus-Martí et al (2018, ISBN:978-84-09-07541-6), Creus-Martí et al (2021) <doi:10.1155/2021/9951817> and Creus-Martí et al (2022) <doi:10.1155/2022/4907527>.

License GPL-3

Encoding UTF-8

LazyData true

Imports xtable, ggplot2, reshape2, compositions, MASS, broom, R2jags, ggbiplot, zCompositions, utils

Suggests testthat (>= 3.0.0), gtools, knitr, parallel, base

Config/testthat/edition 3

RoxygenNote 7.3.2

NeedsCompilation no

Repository CRAN

Date/Publication 2025-06-13 12:00:02 UTC

Contents

CoDaLoMic-package 3
B1MODImatrizP 4
B1MODImodel 5
Balance 6
BPBM_Matrix 6
cockroach 7

Estimate_Param_EstParmFunc 8
Estimate_Param_FBM 9
Estimating_BPBM 11
EstParmFunc 13
EstParmFunc_FBM 14
ExpectedValuess_BPBM 15
ExpectedValues_EstParmFunc_FBM 16
ExpectedValue_EstParmFunc 18
FromVectorToMatrix_BPBM 19
FVectorPBmodeloPredi 20
Graphics 21
GraphicsPrediction 22
GraphicsPredictionBPBM 24
GraphicsSPBal 27
LogVeroFuncBUENA 28
MaxBacteria 29
MaxBacteriaPred 30
ObtainigValuePB 32
ObtainigValueSPBal 33
ObtainingDIC 34
PBalance 36
PBalancePredi 36
PBmodel 37
PCAbiplot 38
Percen_Variance 39
PlotDendogram 40
PredictionBPBM 41
PredictionEstParmFunc 43
PredictionFBM 45
PreparingTheData 47
QualityControl 48
ridgeregression 49
rxnrate 50
Simulated 51
StudyingParam 51
TableBPBM 52
TableFBM 54
Table_alr_Dirich_glv 55
TauAndParameters_EstParmFunc_FBM 56
vecttor 58
ZeroData 59

Index **61**

`CoDaLoMic-package` *CoDaLoMic*

Description

This package contain functions to model compositional and longitudinal microbiome datasets. This package contains the functions needed to execute the models published in the articles:

- Creus-Martí, I., Moya, A., Santonja, F. J. (2018), A Statistical Model with a Lotka-Volterra Structure for Microbiota Data, Lucas Jodar, Juan Carlos Cortes and Luis Acedo, Modelling for engineering and human behavior 2018, Instituto Universitario de Matematica Multidisciplinar, ISBN: 978-84-09-07541-6.
- Creus-Martí, I., Moya, A., Santonja, F. J. (2021). A Dirichlet autoregressive model for the analysis of microbiota time-series data. Complexity, 2021, 1-16.
- Creus-Martí, I., Moya, A., Santonja, F. J. (2022). Bayesian hierarchical compositional models for analysing longitudinal abundance data from microbiome studies. Complexity, 2022.

In addition, the package contains one real dataset extracted from:

- Marín-Miret, J., Pérez-Cobas, A. E., Domínguez-Santos, R., Pérez-Rocher, B., Latorre, A., & Moya, A. (2024). Adaptability of the gut microbiota of the German cockroach Blattella germanica to a periodic antibiotic treatment. Microbiological Research, 287, 127863.

Details

We refer to the model described in Creus-Martí (2018) as Dirich-gLV, we refer to the model described in Creus-Martí (2021) as FBM and we refer to the model described in Creus-Martí (2022) as BPBM.

Access to Additional Files

This package includes files in the directory ‘inst/extdata‘. Users can access these files using ‘system.file()‘. The following files are available:

- `README.pdf`, `README.Rmd`, `README.R`: Basic instructions for using the package.
- `1-s2.0-S0944501324002647-mmc6.xlsx`: Original cockroach dataset extracted from Marín-Miret et al (2024).
- `Simulated.R`: Code use to obtain the Simulated dataset.
- `cockroach.R`: Code use to obtain the cockroach dataset.

To access these files, use: `base::system.file("extdata", "filename", package = "CoDaLoMic")`

For instance: `base::system.file("extdata", "README.pdf", package = "CoDaLoMic")`

On Windows, the PDF can be opened as follows: `base::shell.exec(system.file("extdata", "README.pdf", package = "CoDaLoMic"))`

Author(s)

Maintainer: Irene Creus Martí <`ircrmar@mat.upv.es`> (ORCID)

`B1MODImatrizP` *Ridge regression matrix*

Description

Defining the part of the ridge regression matrix that carries the information of the bacteria `especieII`.

Usage

```
B1MODImatrizP(Tt, especieII, especie, E, EspecieMaxima)
```

Arguments

`Tt`	Number of time points available
`especieII`	Number. The number of the row in which the bacteria that we want to use is placed in the matrix `especie`.
`especie`	Matrix that contains at row i the bacterial taxa of bacteria i at all time points.
`E`	Number of bacteria available
`EspecieMaxima`	Row in which the bacteria chosen as reference is in `especie`. This is the bacteria that is going to be at the denominator of the balance and at the denominator of the alr transformation.

Value

Returns a matrix. The first column contain the number 1 repeated `Tt` times. The second column contains the alr transformation of the `especieII` in all time points. The third column contains the balance (whose numerator has all the bacteria except `especieII` and `EspecieMaxima` and the denominator contains the `EspecieMaxima`) in all time points.

Examples

```
Tt=2
especie1=cbind(c(0.5,0.3,0.2), c(0.1,0.3,0.6))
especieII=1
E=3
EspecieMaxima=3

B1MODImatrizP(Tt, especieII,especie1, E, EspecieMaxima)
```

`B1MODImodel` *Obtaining the regression value of the FBM*

Description

This function calculates the value of the FBM regression, defined by:

$\mu_{it} = a_{i1} + a_{i2} \cdot \text{alr}(x_{i,(t-1)}) + a_{i3} \cdot \text{Balance}(x_{i,(t-1)})$ for $i = 1, \ldots, D-1$ where D is the number of bacteria

Usage

```
B1MODImodel(A, especie, E, EspecieMaxima, Tt)
```

Arguments

`A`	Matrix of dimensions (E-1)x3 that contains all the parameters of the model except tau
`especie`	Matrix that contains at row i the bacterial taxa of bacteria i at all time points.
`E`	Number of bacteria available
`EspecieMaxima`	Row in which the bacteria chosen as reference is in `especie`.This bacteria is used as reference in the alr tranformation that the model does and it is placed at the denominator of the balance)
`Tt`	Number of time points available.

Value

Returns a matrix. The row i contains the regression values of the bacteria i at all time points.

References

Creus-Martí, I., Moya, A., Santonja, F. J. (2021). A Dirichlet autoregressive model for the analysis of microbiota time-series data. Complexity, 2021, 1-16.

Examples

```
df<-data.frame(cbind(c(0.1,0.1,0.8),c(0.2,0.1,0.7)))
E=3
EspecieMaxima=3
set.seed(724)
A=matrix(c(-2:3),2,3)
Tt=2

B1MODImodel(A,df, E, EspecieMaxima,Tt)
```

`Balance` *Calculating the balance of the FBM model*

Description

Defining a balance where we compare all the bacteria (except the one chosen as reference and the `especieI`) with the one chosen as reference.

Usage

```
Balance(A, especieI, especie, E, EspecieMaxima)
```

Arguments

`A`	Number of time points for which we calculate the balance
`especieI`	Number. The bacteria that we do not include in the balance. We must write the number of the row in which the bacteria is placed in the matrix `especie`.
`especie`	Matrix that contains at row i the bacterial taxa of bacteria i at all time points.
`E`	Number of bacteria available
`EspecieMaxima`	Row in which the bacteria chosen as reference is in `especie`.This bacteria is used as reference in the alr tranformation that the model does and it is placed at the denominator of the balance)

Value

Returns a vector with the value of the balance for all the time points indicated.

Examples

```
Balance(2,2,cbind(c(0.1,0.1,0.8),c(0.2,0.1,0.7)),3,3)
```

`BPBM_Matrix` *Obtains the matrix of covariates of the BPBM*

Description

This function writes the matrix of covariates of the BPBM.

Usage

```
BPBM_Matrix(rows.position, PB, Tt)
```

Arguments

`rows.position`	Vector. Vector with the number of the rows where the SPBal are in the matrix PB. We write first the row in which the balance with higher variance is placed, then the row in which the balance with second higher variance is placed...
`PB`	Matrix. Each line os the matrix PB contains the values of one principal balance at all time points.
`Tt`	Number of time points available.

Details

In an example with two SPBal and three time points, the covariates are written in the following order:

$$\begin{array}{rrr} 1 & 1 & 1 \\ SPBal_{1,t-1} & SPBal_{1,t-2} & SPBal_{1,t-3} \\ SPBal_{2,t-1} & SPBal_{2,t-2} & SPBal_{2,t-3} \end{array}$$

Value

Returns a matrix with the covariates of the model.

References

Creus-Martí, I., Moya, A., Santonja, F. J. (2022). Bayesian hierarchical compositional models for analysing longitudinal abundance data from microbiome studies. Complexity, 2022.

Examples

```
matt=matrix(c(1:12),3,4)
rows.position=c(2,3)
BPBM_Matrix(rows.position,matt,4)
```

`cockroach` *cockroach dataset*

Description

Gut microbiome dataset of a Blatella germanica cockroach treated by kanamycin during three periods of time (days: 1–10, 36–45, 71–80). The data is extracted from Marín-Miret et al (2024), more specifically, the data is the information of the K3 cockroach in the article. The dataset contains 105 time points and 210 genera.

Usage

```
data(cockroach)
```

Format

A data frame with 105 rows and 211 columns.

References

Marín-Miret, J., Pérez-Cobas, A. E., Domínguez-Santos, R., Pérez-Rocher, B., Latorre, A., & Moya, A. (2024). Adaptability of the gut microbiota of the German cockroach Blattella germanica to a periodic antibiotic treatment. Microbiological Research, 287, 127863.

`Estimate_Param_EstParmFunc`

Estimating Parameters of EstParmFunc

Description

This function calculates the estimated parameters of the Dirich-gLV model.

Usage

```
Estimate_Param_EstParmFunc(Iter.EstParmFunc, paramini, especie, seed = NULL)
```

Arguments

`Iter.EstParmFunc`	Number. Number of iterations.
`paramini`	Initial values of the parameters. Vector equal to `c(tau.ini, as.vector(pam.ini))` where:
`especie`	Matrix that contains at row i the bacterial taxa of bacteria i at all time points. The bacteria placed in the last row of the matrix will be used as reference in the alr transformation.
`seed`	Number. Set a seed. Default `seed=NULL`.

- pam.ini Matrix. Each row has the parameters of each bacteria, following the same structure than pam in EstParmFunc
- tau.ini Number. Initial value of the tau parameter in the model

Details

Maximum likelihood estimation is used. This function makes an iterative process, it obtains the value of the parameter that maximize the Dirichlet loglikelihood (defined in EstParmFunc) using the Nelder-Mead method and some initial parameters. Then it uses this value as initial parameters and repeats the process `Iter.EstParmFunc` times.

Value

Returns a list with:

- All.iter: Matrix. Each row has the parameters obtained in each iteration. The parameters are in the columns written in the same order that they are written in `paramini`. In this matrix we must observe that in the last iterations the values has really similar or equal values, it not, we need to increase the value of `Iter.EstParmFunc`.
- Param.Estimates: The estimated parameters. The parameters are in the columns written in the same order that they are written in `paramini`.

References

Creus-Martí, I. and Moya, A. and Santonja, F. J. (2018). A Statistical Model with a Lotka-Volterra Structure for Microbiota Data. Lucas Jodar, Juan Carlos Cortes and Luis Acedo, Modelling for engineering and human behavior 2018, Instituto Universitario de Matematica Multidisciplinar. ISBN: 978-84-09-07541-6

Examples

```
especie=cbind(c(0.5,0.3,0.2),c(0.1,0.3,0.6))
paramini=c(100,2,3,4,5,6,7)
Estimate_Param_EstParmFunc(5, paramini , especie,714)
```

`Estimate_Param_FBM` *Estimating Parameters of EstParmFunc_FBM*

Description

This function estimates the parameters of the FBM model.

Usage

```
Estimate_Param_FBM(
  tau,
  ridge.final,
  Iter.EstParmFunc = 80,
  especie,
  EspecieMaxima,
  Tt,
  E,
  seed = NULL
)
```

Arguments

`tau`	Number. Value of the tau parameter.
`ridge.final`	Object of class "ridgelm". Values obtained with the ridge regression.
`Iter.EstParmFunc`	Number. Number of iterations. Default: 80 iterations.
`especie`	Matrix that contains at row i the bacterial taxa of bacteria i at all time points. The bacteria placed in the last row of the matrix will be used as reference in the alr transformation and will be at the denominator of the balance.
`EspecieMaxima`	Row in which the bacteria chosen as reference is in `especie`. This is the bacteria that is going to be at the denominator of the balance and in the deniminator of the alr tranformation. As a result, in this function, `EspecieMaxima` must be equal to `E`.
`Tt`	Number of time points available
`E`	Number. Number of bacteria available.
`seed`	Number. Set a seed. Default `seed=NULL`.

Details

Maximum likelihood estimation is used. This function makes an iterative process, for a given value of tau, it obtains the value of the rest of the parameters that maximize the dirichlet loglikelihood (defined in EstParmFunc_FBM) using the Nelder-Mead method and the values obtained in the ridge regression as initial parameters. Then it uses the values obtained as initial parameters and repeats the process `Iter.EstParmFunc` times.

The regression of this model is defined by

$$\mu_{it} = a_{i1} + a_{i2} \cdot \text{alr}(x_{i,(t-1)}) + a_{i3} \cdot \text{Balance}(x_{i,(t-1)}) \text{ for } i = 1, \ldots, D-1 \text{ where } D \text{ is the number of bacteria}$$

Value

Returns a list with:

- All.iter: Matrix. Each row has the parameters obtained in each iteration. The parameters are in the columns written in following order: a11,a12,a13, a21, a22,a23, ...a(D-1)1,a(D-1)2,a(D-1)3,tau. Where D is the number of bacterial species present in the matrix `especie`. In this matrix we must observe that in the last iterations the values has really similar or equal values, it not, we need to increase the value of `Iter.EstParmFunc`.
- Param.Estimates: Vector with the estimated parameters, in the following order: a11,a12,a13, a21, a22,a23, ...a(D-1)1,a(D-1)2,a(D-1)3,tau. Where D is the number of bacterial species present in the matrix `especie`.
- AIC Number: Value of the AIC.

References

Creus-Martí, I., Moya, A., Santonja, F. J. (2021). A Dirichlet autoregressive model for the analysis of microbiota time-series data. Complexity, 2021, 1-16.

Examples

```
set.seed(123)
especie=t(gtools::rdirichlet(5,c(1,3,1)))
Tt=5
E=3
EspecieMaxima=3
ridge.final=ridgeregression(Tt,especie, E, EspecieMaxima)
tau=20
Iter.EstParmFunc=40

Estimate_Param_FBM(tau,ridge.final,Iter.EstParmFunc, especie,EspecieMaxima,Tt,E, 714)
```

`Estimating_BPBM` *Estimating BPBM*

Description

The estimation of the BPBM model is carried out using MCMC. To execute this function it is necessary to have the program Just Another Gibbs Sampler (JAGS) (Plummer, 2003) program installed.

Usage

```
Estimating_BPBM(
  especie,
  Tt,
  E,
  MatrizPBmodelo,
  nn.chain = 3,
  nn.burnin = 5000,
  nn.sample = 20000,
  nn.thin = 10,
  seed = NULL
)
```

Arguments

`especie`	Matrix that contains at row i the bacterial taxa of bacteria i at all time points.
`Tt`	Number of time points available.
`E`	Number of bacteria in the dataset.
`MatrizPBmodelo`	Matrix with the covariates of the model. In an example with two SPBal and three time points, the covariates are written in the following order:

$$\begin{array}{ccc} 1 & 1 & 1 \\ SPBal_{1,t-1} & SPBal_{1,t-2} & SPBal_{1,t-3} \\ SPBal_{2,t-1} & SPBal_{2,t-2} & SPBal_{2,t-3} \end{array}$$

`nn.chain`	the number of chains to use with the simulation. Default is 3, minimum2.
`nn.burnin`	the number of burnin iterations. Default is 5000.
`nn.sample`	the number of iterations to take. Default: 20000. The markov chain will have ("sample"-"burnin")/"thin" iterations.
`nn.thin`	the thinning interval to be used. Default: 10.
`seed`	Number. Set a seed. Default `seed=NULL`.

Value

Returns a list with:

- List with:
  - R2jagsOutput: R2jags object with the information of the estimation.
  - SamplesAllChains: Matrix. Matrix that has the iterations of all the Markov chains joined.

References

- Creus-Martí, I., Moya, A., Santonja, F. J. (2022). Bayesian hierarchical compositional models for analysing longitudinal abundance data from microbiome studies. Complexity, 2022.
- Plummer, M. (2003, March). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In Proceedings of the 3rd international workshop on distributed statistical computing (Vol. 124, No. 125.10, pp. 1-10).

Examples

```
set.seed(314)
especie=t(gtools::rdirichlet(n=6, c(6,6,1,6,6)))
E=5
Tt=6
MatrizPBmodelo=rbind(c(1,1,1,1,1,1),c(-0.3,0.4,0.3,-0.7,-0.4,-0.6),c(0.3,0.5,-0.3,0.1,0.4,0.1))



Estimating_BPBM(especie,
               Tt,
               E,
               MatrizPBmodelo,
               nn.chain=3,
               nn.burnin=1000,
               nn.sample=2000,
               nn.thin=10,
               714)
```

`EstParmFunc` *Writting the loglikelihood of the dirichlet*

Description

This function calculates the loglikelihood of the dirichlet for the Dirich-gLV model.

Usage

```
EstParmFunc(parms.vector, especie)
```

Arguments

`parms.vector`	Vector equal to `c(tau,as.vector( pam))` where: - pam: Matrix. Each row has the parameters of each bacteria. Following our example, pam has the parameters placed as follows: r1 a11 a12 r2 a21 a22 - tau: Number. Value of the tau parameter in the model
`especie`	Matrix that contains at row i the bacterial taxa of bacteria i at all time points . The bacteria placed in the last row of this matrix is the one used as reference in the alr transfromation that the model apply

Details

In an example with three bacteria, the regression of this model is defined by

$$r_1 \cdot log(x_1(t)/x_3(t)) + log(x_1(t)/x_3(t)) \cdot [a_{11} \cdot log(x_1(t)/x_3(t))(t) + a_{12} \cdot log(x_2(t)/x_3(t))]$$

$$r_2 \cdot log(x_2(t)/x_3(t)) + log(x_2(t)/x_3(t)) \cdot [a_{21} \cdot log(x_1(t)/x_3(t))(t) + a_{22} \cdot log(x_2(t)/x_3(t))]$$

Value

Returns a number with the value of the dirichlet loglikelihood.

References

Creus-Martí, I. and Moya, A. and Santonja, F. J. (2018). A Statistical Model with a Lotka-Volterra Structure for Microbiota Data. Lucas Jodar, Juan Carlos Cortes and Luis Acedo, Modelling or engineering and human behavior 2018, Instituto Universitario de Matematica Multidisciplinar. ISBN: 978-84-09-07541-6

Examples

```
especie1=cbind(c(0.5,0.3,0.2), c(0.1,0.3,0.6))
tau1=0.4
parms1= cbind(c(0.1,0.2),c(-0.2,0.1),c(0.3,0.2))
parms11=c(tau1,as.vector( parms1))

EstParmFunc(parms11,especie1)
```

`EstParmFunc_FBM` *Writting the loglikelihood of the dirichlet*

Description

This function calculates the loglikelihood of the dirichlet for the FBM model.

Usage

```
EstParmFunc_FBM(param, especie, E, EspecieMaxima, Tt, especiemodi)
```

Arguments

`param`	Vector with the parameters in the following order: a11,a12,a13, a21, a22,a23, ...a(D-1)1,a(D-1)2,a(D-1)3,tau. Where D is the number of bacterial species present in the matrix `especie`.
`especie`	Matrix that contains at row i the bacterial taxa of bacteria i at all time points.
`E`	Number of bacteria available
`EspecieMaxima`	Row in which the bacterial with maximum mean abundance is in `especie`.This bacteria is used as reference in the alr transformation that the model does and it is placed at the denominator of the balance)
`Tt`	Number of bacteria available
`especiemodi`	Matrix that contains at row i the bacterial taxa of bacteria i at time points t=2,...,`Tt`. The bacteria are placed in the same order than in `especie`.

Details

The regression of this model is defined by

$$\mu_{it} = a_{i1} + a_{i2} \cdot \text{alr}(x_{i,(t-1)}) + a_{i3} \cdot \text{Balance}(x_{i,(t-1)}) \text{ for } i = 1, \ldots, D-1 \text{ where } D \text{ is the number of bacteria}$$

Value

Returns a number with the value of the dirichlet loglikelihood.

References

Creus-Martí, I., Moya, A., Santonja, F. J. (2021). A Dirichlet autoregressive model for the analysis of microbiota time-series data. Complexity, 2021, 1-16.

Examples

```
especie1=cbind(c(0.5,0.3,0.2), c(0.1,0.3,0.6))
especiemodi=especie1[,-1]
tau1=0.4
parms1= cbind(c(0.1,0.2),c(-0.2,0.1),c(0.3,0.2))
parms11=c(as.vector( t(parms1)),tau1)

EstParmFunc_FBM(parms11,especie1,3 ,3  , 2,especiemodi)
```

`ExpectedValuess_BPBM` *Obtainig the value of the dirichlet parameters, the expected value and the variance.*

Description

This function calculates the value of the dirichlet parameters, the expected value and the variance for the BPBM model.

Usage

```
ExpectedValuess_BPBM(Estimated.Param, MatrizPBmodelo, E, Tt)
```

Arguments

`Estimated.Param`	Vector with the estimate parameters. Column "mean" of the output of "StudyingParam" function.
`MatrizPBmodelo`	Matrix. Output of "ObtainingValueSPBal" called "MatrixSPBal".
`E`	Number of bacteria available.
`Tt`	Number of time points available.

Details

The regression of this model is defined by:

$$\mu_{it} = a_{i0} + a_{i1} \cdot \text{SPBal}_{1,t-1} + \cdots + a_{iM} \cdot \text{SPBal}_{M,t-1}$$

Value

Returns a list with:

- Dirichlet.Param: Matrix. Matrix that contains at row i the dirichlet parameter of the bacteria i at all time points.
- Expected.Value: Matrix. Matrix that contains at row i the expected value of the bacteria i at all time points. The bacterias are placed at the same orden than in `especies`.
- Variance.Value: Matrix. Matrix that contains at row i the variance of the bacteria i at all time points. The bacterias are placed at the same orden than in `especies`.

References

Creus-Martí, I., Moya, A., Santonja, F. J. (2022). Bayesian hierarchical compositional models for analysing longitudinal abundance data from microbiome studies. Complexity, 2022.

Examples

```
Tt=3
E=3
Estimated.Param=c(0.1 ,0.4, 0.7, 0.2 ,0.5, 0.8 ,0.3, 0.6, 0.9,
                0.1, 0.4 ,0.7, 0.2, 0.5, 0.8, 0.3 ,0.6, 0.9)
MatrizPBmodelo=rbind(c(1,1,1),c(0.3,0.6,-0.1),c(0.2,-0.4,0.3))
ExpectedValuess_BPBM(Estimated.Param,MatrizPBmodelo,E,Tt)
```

`ExpectedValues_EstParmFunc_FBM`

Obtainig the value of the dirichlet parameters, the expected value and the variance.

Description

This function calculates the value of the dirichlet parameters, the expected value and the variance for the FBM model.

Usage

```
ExpectedValues_EstParmFunc_FBM(
  paramEstimadosFinal,
  especie,
  E,
  EspecieMaxima,
  Tt
)
```

Arguments

`paramEstimadosFinal`	The estimate parameters, in the following order: a11,a12,a13, a21, a22,a23, ...a(D-1)1,a(D-1)2,a(D-1)3,tau. Where D is the number of bacterial species present in the matrix `especie`.
`especie`	Matrix that contains at row i the bacterial taxa of bacteria i at all time points.
`E`	Number of bacteria available.
`EspecieMaxima`	Row in which the bacteria chosen as reference is in `especie`.This bacteria is used as reference in the alr tranformation that the model does and it is placed at the denominator of the balance).
`Tt`	Number of time points available.

Details

The regression of this model is defined by

$$\mu_{it} = a_{i1} + a_{i2} \cdot \text{alr}(x_{i,(t-1)}) + a_{i3} \cdot \text{Balance}(x_{i,(t-1)}) \text{ for } i = 1, \ldots, D-1 \text{ where } D \text{ is the number of bacteria}$$

Value

Returns a list with:

- Dirichlet.Param: Matrix. Matrix that contains at row i the dirichlet parameter of the bacteria i at all time points.
- Expected.Value: Matrix. Matrix that contains at row i the expected value of the bacteria i at all time points. The bacterias are placed at the same orden than in `especies`.
- Variance.Value: Matrix. Matrix that contains at row i the variance of the bacteria i at all time points. The bacterias are placed at the same orden than in `especies`.

References

Creus-Martí, I., Moya, A., Santonja, F. J. (2021). A Dirichlet autoregressive model for the analysis of microbiota time-series data. Complexity, 2021, 1-16.

Examples

```
set.seed(123)
especie=t(gtools::rdirichlet(2,c(1,1,3)))
Tt=2
E=3
tau=5
EspecieMaxima=3
Iter.EstParmFunc=5
parms11=c(0.1,0.2,0.3,0.4,0.5,0.6,tau)

ExpectedValues_EstParmFunc_FBM(parms11 , especie,E,EspecieMaxima,Tt)
```

`ExpectedValue_EstParmFunc` *Obtaining the value of the Dirichlet parameters, the expected value and the variance.*

Description

This function calculates the value of the Dirichlet parameters, the expected value and the variance for the Dirich-gLV model.

Usage

```
ExpectedValue_EstParmFunc(Param.Estimates, especie)
```

Arguments

`Param.Estimates`	Vector with the estimated parameters. This value is the output of the Estimate_Param_EstParmFunc function.
`especie`	Matrix that contains at row i the bacterial taxa of bacteria i at all time points. The bacteria placed in the last row of this matrix is the one used as reference in the alr transformation that the model applies.

Value

Returns a list with:

- Dirichlet.Param: Matrix. Matrix that contains at row i the Dirichlet parameter of the bacteria i at all time points.
- Expected.Value: Matrix. Matrix that contains at row i the expected value of the bacteria i at all time points. The bacterias are placed at the same orden than in `especies`.
- Variance.Value: Matrix. Matrix that contains at row i the variance of the bacteria i at all time points. The bacterias are placed at the same orden than in `especies`.

References

Creus-Martí, I. and Moya, A. and Santonja, F. J. (2018). A Statistical Model with a Lotka-Volterra Structure for Microbiota Data. Lucas Jodar, Juan Carlos Cortes and Luis Acedo, Modelling or engineering and human behavior 2018, Instituto Universitario de Matematica Multidisciplinar. ISBN: 978-84-09-07541-6

Examples

```
especie1=cbind(c(0.5,0.3,0.2), c(0.1,0.3,0.6))
tau1=0.4
parms1= cbind(c(0.1,0.2),c(-0.2,0.1),c(0.3,0.2))
```

```
parms11=c(tau1,as.vector( parms1))

ExpectedValue_EstParmFunc(parms11,especie1)
```

`FromVectorToMatrix_BPBM`

Writting the parameters in the matrix form required in BPBM model

---

Description

`StudyingParam` returns a matrix where the value of the parameters is in the column "mean". This function inputs this columns and outputs the parameters in the matrix form required by the BPBM model.

Usage

```
FromVectorToMatrix_BPBM(param, MatrizPBmodelo, E)
```

Arguments

`param`	Vector. Column "mean" of the output of "StudyingParam" function.
`MatrizPBmodelo`	Matrix with the covariates of the model. In an example with two SPBal and three time points, the covariates are written in the following order: $\begin{array}{ccc} 1 & 1 & 1 \\ SPBal_{1,t-1} & SPBal_{1,t-2} & SPBal_{1,t-3} \\ SPBal_{2,t-1} & SPBal_{2,t-2} & SPBal_{2,t-3} \end{array}$
`E`	Number of bacteria in the dataset.

Value

Returns a matrix with the parameters in the order required by the BPBM model.

References

Creus-Martí, I., Moya, A., Santonja, F. J. (2022). Bayesian hierarchical compositional models for analysing longitudinal abundance data from microbiome studies. Complexity, 2022.

Examples

```
set.seed(314)
especie=t(gtools::rdirichlet(n=6, c(6,6,1,6,6)))
E=5
Tt=6
MatrizPBmodelo=rbind(c(1,1,1,1,1,1),
                     c(-0.3,0.4,0.3,-0.7,-0.4,-0.6),
                     c(0.3,0.5,-0.3,0.1,0.4,0.1))
```

```
est=Estimating_BPBM(especie,
                   Tt,
                   E,
                   MatrizPBmodelo,
                   nn.chain=3,
                   nn.burnin=1000,
                   nn.sample=5000,
                   nn.thin=10)

mm=StudyingParam(est$R2jagsOutput$BUGSoutput$summary,est$SamplesAllChains)
FromVectorToMatrix_BPBM(mm$Param.Summary[,"mean"],MatrizPBmodelo,5)
```

`FVectorPBmodeloPredi` *Obtaining a vector with the covariates of the prediction*

Description

Calculates a vector with the covariates of the BPBM model in one time point.

Usage

```
FVectorPBmodeloPredi(NumSPBal, DemSPBal, v, MatrizPBmodelo)
```

Arguments

`NumSPBal`	List. The component i of the list has the number of the row of the matrix `especie` where the bacteria in the numerator of the principal balance i are placed.
`DemSPBal`	List. The component i of the list has the number of the row of the matrix `especie` where the bacteria in the denominator of the principal balance i are placed.
`v`	Vector. Vector with a coda composition. The bacteria ar in the same orden than the matrix `especie`
`MatrizPBmodelo`	the matrix that contains the covariates of the model. The first line es equal to 1 for all columns. The other rows contain the value of one SPBal at all time points. The selected principal balance of the row i+1 has at its numerator the bacteria placed in the rows `NumSPBal[[i]]` of the "especie". The selected principal balance of the row i+1 has at its denominator the bacteria placed in the rows `DemSPBal[[i]]` of the "especie".

Value

Returns a vector where the first component is a 1 and the following components have the values of the SPBal. The SPBal in the component i+1 has at its numerator the bacteria placed in the rows `Num[[i]]` of the `especie`. The SPBal of the component i+1 has at its denominator the bacteria placed in the rows `Dem[[i]]` of the `especie`.

References

Creus-Martí, I., Moya, A., Santonja, F. J. (2022). Bayesian hierarchical compositional models for analysing longitudinal abundance data from microbiome studies. Complexity, 2022.

Examples

```
v=c(0.1,0.1,0.2,0.3,0.3)
Num2<-list(3,c(3,5),1,c(3,5,4))
Dem2<-list(5,4,2,c(1,2))
MatrizPBmodelo=rbind(c(1,1,1,1),
                   c(-0.3,0.2,0.5,0.6),
                   c(-0.4,0.3,0.5,0.6),
                   c(0.5,0.3,0.2,0.7),
                   c(-0.2,0.9,0.2,0.1)   )

FVectorPBmodeloPredi(Num2,Dem2,v,MatrizPBmodelo)
```

`Graphics` *Plots the time series*

Description

Plots the time series

Usage

```
Graphics(especie, names.especie, esperanza, Variance, Plot.Tipe, Detail)
```

Arguments

`especie`	Matrix that contains at row i the bacterial taxa of bacteria i at all time points.
`names.especie`	Vector with the names of the bacteria in the same order that are placed in the `especie` matrix.
`esperanza`	Matrix that contains at row i the expected value of the bacterial taxa i at all time points. The bacteria must be placed in the same order than in `especie`
`Variance`	Matrix that contains at row i the variance of the bacterial taxa i at all time points. The bacteria must be placed in the same order than in `especie`
`Plot.Tipe`	Character. If `Plot.Tipe==Data` the function displays a graphic of the dataset, if `Plot.Tipe==DataExpected` the function displays a graphic of the data and the expected values, if `Plot.Tipe==All` the function displais a graphic woth the data, the expected values and the variance.If `Plot.Tipe==Var` the function returns the boxplot of the variance at each time point and the variance of each bacteria. If `Plot.Tipe==OnlyVar` the function returns the boxplots of the variance at each time points.
`Detail`	Character. If `Detail==no` the graphic obtained when `Plot.Tipe==DataExpected` and `Plot.Tipe==All` will have the same y axis for all the taxa. If `Detail==yes` these functions will have different y axis.

Value

Returns the indicated plots.

Examples

```
names.especie=c("Bact1", "Bact2", "Bact3")
especie=cbind(c(0.5,0.3,0.2), c(0.6,0.3,0.1),c(0.4,0.1,0.5),c(0.4,0.1,0.5))
esperanza=especie[,c(1:3)]+0.1
Variance=matrix(c(runif(9,0.001,0.004)), 3,3)

Graphics(especie, names.especie, esperanza, Variance,"Data","no")
Graphics(especie, names.especie, esperanza, Variance,"DataExpected","no")
Graphics(especie, names.especie, esperanza, Variance,"All","no")
```

`GraphicsPrediction` *Plots the time series*

Description

This function takes into account the data used to estimate and the data used to predict.

Usage

```
GraphicsPrediction(
  especie.All,
  names.especie,
  ExpectedValue.All,
  VarianceValue.All,
  Pred,
  Plot.Tipe,
  Detail
)
```

Arguments

`especie.All`	Matrix that contains at row i the bacterial taxa of bacteria i at all time points.
`names.especie`	Vector with the names of the bacteria in the same order that are placed in the `especie` matrix.
`ExpectedValue.All`	Matrix that contains at row i the expected value of the bacterial taxa i at all time points. The bacteria must be placed in the same order than in `especie`. This matrix must comply: dim(ExpectedValue.All)[2]=dim(especie.All)[2]-1
`VarianceValue.All`	Matrix that contains at row i the variance of the bacterial taxa i at all time points. The bacteria must be placed in the same order than in `especie`.This matrix must comply: dim(VarianceValue.All)[2]=dim(especie.All)[2]-1

`Pred`	Number. Indicates the time point in which we start predicting.
`Plot.Tipe`	Character. If `Plot.Tipe==Data` the function displays a graphic of the dataset, if `Plot.Tipe==DataExpected` the function displays a graphic of the data and the expected values, if `Plot.Tipe==All` the function displays a graphic with the data, the expected values and the variance.If `Plot.Tipe==Var` the function returns the boxplot of the variance at each time point and the variance of each bacteria. If `Plot.Tipe==OnlyVar` the function returns the boxplots of the variance at each time points.
`Detail`	Character. If `Detail==no` the graphic obtained when `Plot.Tipe==DataExpected` and `Plot.Tipe==All` will have the same y axis for all the taxa. If `Detail==yes` these functions will have different y axis.

Value

Returns the indicated plots with a vertical line when the time point is equal to Pred-1, in Pred the prediction has started.

Examples

```
names.especie=c("Bact1", "Bact2", "Bact3")
especie.All=cbind(c(0.5,0.3,0.2),
                c(0.6,0.3,0.1),
                c(0.4,0.1,0.5),
                c(0.4,0.1,0.5),
                c(0.4,0.1,0.5),
                c(0.4,0.1,0.5))
ExpectedValue.All=especie.All[,-1]+0.1
VarianceValue.All=matrix(c(runif(15,0.001,0.004)), 3,5)
Pred=4

GraphicsPrediction(especie.All,
                  names.especie,
                  ExpectedValue.All,
                  VarianceValue.All ,
                  Pred,
                  "Data",
                  "no")

GraphicsPrediction(especie.All,
                 names.especie,
                 ExpectedValue.All,
                 VarianceValue.All ,
                 Pred,
                 "DataExpected",
                 "no")

GraphicsPrediction(especie.All,
                 names.especie,
                 ExpectedValue.All,
                 VarianceValue.All ,
```

```
                 Pred,
                 "All",
                 "no")

GraphicsPrediction(especie.All,
                 names.especie,
                 ExpectedValue.All,
                 VarianceValue.All ,
                 Pred,
                 "Var",
                 "no")

GraphicsPrediction(especie.All,
                 names.especie,
                 ExpectedValue.All,
                 VarianceValue.All ,
                 Pred,
                 "OnlyVar",
                 "no")
```

`GraphicsPredictionBPBM` *Plots the time series*

Description

This function takes into account the data used to estimate and the data used to predict. We use this function when we want to observe the results obtained with the BPBM model.

Usage

```
GraphicsPredictionBPBM(
  especie.All,
  names.especie,
  ExpectedValue.All,
  VarianceValue.All,
  Pred,
  Plot.Tipe,
  Varmas,
  Varmenos,
  Detail
)
```

Arguments

`especie.All`	Matrix that contains at row i the bacterial taxa of bacteria i at all time points.

`names.especie`	Vector with the names of the bacteria in the same order that are placed in the `especie` matrix.
`ExpectedValue.All`	Matrix that contains at row i the expected value of the bacterial taxa i at all time points. The bacteria must be placed in the same order than in `especie`. This matrix must comply: dim(ExpectedValue.All)[2]=dim(especie.All)[2]-1
`VarianceValue.All`	Matrix that contains at row i the variance of the bacterial taxa i at all time points. The bacteria must be placed in the same order than in `especie`.This matrix must comply: dim(VarianceValue.All)[2]=dim(especie.All)[2]-1
`Pred`	Number. Indicates the time point in which we start predicting.
`Plot.Tipe`	Character. If `Plot.Tipe==Data` the function displays a graphic of the dataset, if `Plot.Tipe==DataExpected` the function displays a graphic of the data and the expected values, if `Plot.Tipe==All` the function displays a graphic with the data, the expected values and the variance (the varmas and varmenos parameters are introduce and are taken into account to plot the variance of the predicted part).If `Plot.Tipe==Var` the function returns the boxplot of the variance at each time point and the variance of each bacteria. If `Plot.Tipe==OnlyVar` the function returns the boxplots of the variance at each time points.
`Varmas`	Matrix. Output of "PredictionBPBM" adding "$ExpVarmas". Matrix that contains at row i the expected value plus two times the sqrt(variance) of the bacteria i at all time points t=`Tt`,...,K, the rest of the time points has 0 values. The bacteria are placed at the same order than in `especies`.
`Varmenos`	Matrix. Output of "PredictionBPBM" adding "$ExpVarmenos". Matrix that contains at row i the expected value minus two times the sqrt(variance) of the bacteria i at all time points t=`Tt`,...,K, the rest of the time points has 0 values. The bacteria are placed at the same order than in `especies`.
`Detail`	Character. If `Detail==no` the graphic obtained when `Plot.Tipe==DataExpected` and `Plot.Tipe==All` will have the same y axis for all the taxa. If `Detail==yes` these functions will have different y axis.

Value

Returns the indicated plots with a vertical line when the time point is equal to `Tt`=Pred-1, in Pred=`Tt`+1 the predicction has started.

References

Creus-Martí, I., Moya, A., Santonja, F. J. (2022). Bayesian hierarchical compositional models for analysing longitudinal abundance data from microbiome studies. Complexity, 2022.

Examples

```
names.especie=c("Bact1", "Bact2", "Bact3")
especie.All=cbind(c(0.5,0.3,0.2),
                c(0.6,0.3,0.1),
                c(0.4,0.1,0.5),
```

```
                 c(0.4,0.1,0.5),
                 c(0.4,0.1,0.5),
                 c(0.4,0.1,0.5))
ExpectedValue.All=especie.All[,-1]+0.1
VarianceValue.All=matrix(c(runif(15,0.001,0.004)), 3,5)
Pred=4
Varmas=cbind(matrix(0,3,2),matrix(c(runif(9,0.001,0.004)) ,3 ,3 ))
Varmenos=cbind(matrix(0,3,2),matrix(c(runif(9,0.001,0.004)) ,3 ,3 ))

GraphicsPredictionBPBM(especie.All,
                      names.especie,
                      ExpectedValue.All,
                      VarianceValue.All ,
                      Pred ,
                      "Data",
                      Varmas,
                      Varmenos,
                      "no")

GraphicsPredictionBPBM(especie.All,
                      names.especie,
                      ExpectedValue.All,
                      VarianceValue.All ,
                      Pred ,
                      "DataExpected",
                      Varmas,
                      Varmenos,
                      "no")

GraphicsPredictionBPBM(especie.All,
                      names.especie,
                      ExpectedValue.All,
                      VarianceValue.All,
                      Pred ,
                      "All",
                      Varmas,
                      Varmenos,
                      "no")

GraphicsPredictionBPBM(especie.All,
                      names.especie,
                      ExpectedValue.All,
                      VarianceValue.All ,
                      Pred ,
                      "Var",
                      Varmas,
                      Varmenos,
                      "no")

GraphicsPredictionBPBM(especie.All,
                      names.especie,
                      ExpectedValue.All,
                      VarianceValue.All ,
```

```
Pred ,
"OnlyVar",
Varmas,
Varmenos,
"no")
```

`GraphicsSPBal` *Obtaining the graphic of the SPBal at all time points*

Description

The SPBal (of BPBM model) are ordered from highest to lowest variance percentage. The zero is highlight because the closer the value of the balance is to zero, the more similar (in terms of relative abundance) the numerator and denominator groups will be. The farther away from zero, the more different.

Usage

```
GraphicsSPBal(MatrizPBmodelo)
```

Arguments

`MatrizPBmodelo`	Matrix. Output of "ObtainigValueSPBal" function. MatrixSPBal is the matrix that contains the covariates of the model. The first line es equal to 1 for all columns. The other rows contain the value of one SPBal at all time points. The selected principal balance of the row i+1 has at its numerator the bacteria placed in the rows `NumSPBal[[i]]` of the "especie". The selected principal balance of the row i+1 has at its denominator the bacteria placed in the rows `DemSPBal[[i]]` of the "especie".

Value

Returns a graphic.

References

Creus-Martí, I., Moya, A., Santonja, F. J. (2022). Bayesian hierarchical compositional models for analysing longitudinal abundance data from microbiome studies. Complexity, 2022.

Examples

```
MatrizPBmodelo=rbind(c(1,1,1,1,1,1),c(-0.3,0.4,0.3,-0.7,-0.4,-0.6),c(0.3,0.5,-0.3,0.1,0.4,0.1))
GraphicsSPBal(MatrizPBmodelo)
```

`LogVeroFuncBUENA` *Writting the loglikelihood of the dirichlet*

Description

This function calculates the loglikelihood of the dirichlet for the BPBM model.

Usage

```
LogVeroFuncBUENA(param, MatrizPBmodelo, E, Tt, especiemodi)
```

Arguments

`param`	Vector. Column "mean" of the output of "StudyingParam" function.
`MatrizPBmodelo`	Matrix with the covariates of the model. In an example with two SPBal and three time points, the covariates are written in the following order: $$\begin{array}{rrr} 1 & 1 & 1 \\ SPBal_{1,t-1} & SPBal_{1,t-2} & SPBal_{1,t-3} \\ SPBal_{2,t-1} & SPBal_{2,t-2} & SPBal_{2,t-3} \end{array}$$
`E`	Number f bacteria in the dataset.
`Tt`	Number of time points available
`especiemodi`	Matrix that contains at row i the bacterial taxa of bacteria i at time points t=2,...,`Tt`.

Value

Returns a number with the value of the dirichlet loglikelihood.

References

Creus-Martí, I., Moya, A., Santonja, F. J. (2022). Bayesian hierarchical compositional models for analysing longitudinal abundance data from microbiome studies. Complexity, 2022.

Examples

```
set.seed(314)
especie=t(gtools::rdirichlet(n=2, c(1,2,3)))
E=3
Tt=2
MatrizPBmodelo=rbind(c(1,1),c(-0.3,0.4),c(0.3,0.5))
set.seed(314)
est=Estimating_BPBM(especie,
                   Tt,
                   E,
                   MatrizPBmodelo,
                   nn.chain=3,
```

```
                 nn.burnin=1000,
                 nn.sample=5000,
                 nn.thin=10)

param=est$SamplesAllChains

especiemodi=especie[,-1]


LogVeroFuncBUENA(param,MatrizPBmodelo,E,Tt,especiemodi)
```

`MaxBacteria` *Putting the reference bacteria at the last row*

Description

This function calculates the mean abundance of each bacteria. Then, it creates a matrix where each row contains the abundance of one bacteria at all time points but the bacteria with maximum (or minimum or a bacteria indicated by the user) mean abundance is placed at the last row

Usage

```
MaxBacteria(nombresOriginal, especieOriginal, E, Tt, which.esp)
```

Arguments

`nombresOriginal`	Vector with the bacterial names at the same order than in DaTa. it must be fulfilled that lenght(nombresOriginal)==dim(DaTa)[2]-1
`especieOriginal`	Matrix that contains at row i the bacterial taxa of bacteria i at all time points
`E`	Number of bacteria available
`Tt`	Number of bacteria available
`which.esp`	If `which=="Max"` this function puts in the last position of the matrix the bacteria with maximum mean abundance. If `which=="Min"` this function puts in the last position of the matrix the bacteria with minimum mean abundance. If which is equal to a number this function puts in the last position of the matrix the bacteria that is in the "which" row of the `especieOriginal` matrix.

Value

Returns a list with

- `especie` - Matrix that contains at row i the bacterial taxa of bacteria i at all time points but the bacteria with maximum (or minimum) mean abundance (or the bacteria indicated by the user) is placed at the last row.

- `especiemodi` - Matrix that contains at row i the bacterial taxa of bacteria i at time points t=2,...,`Tt` but the bacteria with maximum (or minimum) mean abundance (or the bacteria indicated by the user) is placed at the last row.
- `nombres` - Vector with the bacteria's names placed in the order in which appear in the rows of the matrices `especie` and `especiemodi`
- `EE` - Row in which the bacterial with maximum (or minimum) mean abundance was (or the value of "which" if which is numerical).
- `EspecieMaxima` - Row in which the bacterial with maximum (or minimum) mean abundance (or the bacteria indicated by the user) is in `especie`.) #'

Examples

```
names1=c("Bact1","Bact2","Bact3")
set.seed(314)
esp1=t(gtools::rdirichlet(n=4, c(1,3,1)))
e1=3
t1=4

MaxBacteria(names1,esp1,e1,t1,"Max")


names3=c("Bact1","Bact2","Bact3","Bact4","Bact5")
set.seed(314)
esp3=t(gtools::rdirichlet(n=6, c(6,6,1,6,6)))
e3=5
t3=6
MaxBacteria(names3,esp3,e3,t3,"Min")
```

`MaxBacteriaPred` *Putting the reference bacteria at the last row*

Description

This function calculates the mean abundance of each bacteria taking into account the time points used to estimate the model (t=1,2,...,`Tt`). Then, it creates a matrix where each row contains the abundance of one bacteria at all time points but the bacteria with maximum (or minimum) mean abundance (or the bacterial indicated by the user) is placed at the last row

Usage

```
MaxBacteriaPred(
  nombresOriginal,
  especieOriginal,
  E,
  Tt,
  Pred,
```

```
  K,
  especieOriginal.All,
  which.esp
)
```

Arguments

`nombresOriginal`	Vector with the bacterial names at the same order than in DaTa. it must be fulfilled that lenght(nombresOriginal)==dim(DaTa)[2]-1
`especieOriginal`	Matrix that contains at row i the bacterial taxa of bacteria i at t=1,2,...,`Tt`, with `Tt`=Pred-1.
`E`	Number of bacteria available
`Tt`	Number of time points used to estimate the model (`Tt`=Pred-1)
`Pred`	Number. The data at t=1,...,Pred-1 will be used to estimate the model. The rest of the time points will be used to study the capacity of the model to predict. If `Pred==0` all the datatset will be used to estimate the model.
`K`	Number of time points at the data
`especieOriginal.All`	Matrix that contains at row i the bacterial taxa of bacteria i at all time points.
`which.esp`	If `which=="Max"` this function puts in the last position of the matrix the bacteria with maximum mean abundance. If `which=="Min"` this function puts in the last position of the matrix the bacteria with minimum mean abundance. If which is equal to a number this function puts in the last position of the matrix the bacteria that is in the "which" row of the `especieOriginal` matrix.

Value

Returns a list with

- `especie` - Matrix that contains at row i the bacterial taxa of bacteria i at time points t=1,2,...,`Tt` but the bacteria with with maximum (or minimum) mean abundance (or the bacteria indicated by the user) is placed at the last row.
- `especiemodi` - Matrix that contains at row i the bacterial taxa of bacteria i at time points t=2,...,`Tt` but the bacteria with with maximum (or minimum) mean abundance (or the bacteria indicated by the user) is placed at the last row.
- `nombres` - Vector with the bacteria's names placed in the order in which appear in the rows of the matrices `especie` and `especiemodi`
- `EE` - Row in which the bacterial with maximum (or minimum) mean abundance was (or the value of "which" if which is numerical).
- `EspecieMaxima` - Row in which the bacterial with with maximum (or minimum) mean abundance (or the bacteria indicated by the user) is in `especie`.) #' #'
- `especie.all` - Matrix that contains at row i the bacterial taxa of bacteria i at all time points (t=1,2,...,K) but the bacteria with with maximum (or minimum) mean abundance (or the bacteria indicated by the user) is placed at the last row.

- `especiemodi.all` - Matrix that contains at row i the bacterial taxa of bacteria i at all time points (t=2,...,K) but the bacteria with with maximum (or minimum) mean abundance (or the bacteria indicated by the user) is placed at the last row.

Examples

```
names2=c("Bact1","Bact2","Bact3","Bact4","Bact5")
set.seed(314)
esp2=t(gtools::rdirichlet(n=6, c(1,1,5,1,1)))
e2=5

MaxBacteriaPred(names2,esp2[,-c(4,5,6)],e2,3,Pred=4, 6,esp2, "Max")
```

`ObtainigValuePB` *Obtaining the principal balances values*

Description

Calculates the value of the principal balances (Martín-Fernández et al, 2018) at all time points.

Usage

```
ObtainigValuePB(Num, Dem, especie, Tt)
```

Arguments

`Num`	List. The component i of the list has the number of the row of the matrix `especie` where the bacteria in the numerator of the principal balance i are placed.
`Dem`	List. The component i of the list has the number of the row of the matrix `especie` where the bacteria in the denominator of the principal balance i are placed.
`especie`	Matrix that contains at row i the bacterial taxa of bacteria i at all time points.
`Tt`	Number of time points available

Value

Returns a matrix where the row i has the value of the principal balance at all time points. The principal balance of the row i has at its numerator the bacteria placed in the rows `Num[[i]]` of the `especie`. The principal balance of the row i has at its denominator the bacteria placed in the rows `Dem[[i]]` of the `especie`.

References

- Creus-Martí, I., Moya, A., Santonja, F. J. (2022). Bayesian hierarchical compositional models for analysing longitudinal abundance data from microbiome studies. Complexity, 2022.
- Martín-Fernández, J. A., Pawlowsky-Glahn, V., Egozcue, J. J., & Tolosona-Delgado, R. (2018). Advances in principal balances for compositional data. Mathematical Geosciences, 50, 273-298.

Examples

```
set.seed(314)
esp2=t(gtools::rdirichlet(n=6, c(1,1,5,1,1)))

Num2<-list(3,c(3,5),1,c(3,5,4))
Dem2<-list(5,4,2,c(1,2))

ObtainigValuePB(Num2,Dem2,esp2,6)
```

`ObtainigValueSPBal` *Obtaining the selected principal balances values*

Description

Calculates the value of the selected principal balances (SPBal) of the BPBM model at all time points.

Usage

```
ObtainigValueSPBal(Num, Dem, especie, Tt)
```

Arguments

`Num`	List. The component i of the list has the number of the row of the matrix `especie` where the bacteria in the numerator of the principal balance i are placed.
`Dem`	List. The component i of the list has the number of the row of the matrix `especie` where the bacteria in the denominator of the principal balance i are placed.
`especie`	Matrix that contains at row i the bacterial taxa of bacteria i at all time points.
`Tt`	Number of time points available

Value

Returns a list with:

- NumSPBal: List. The component i of the list has the number of the row of the matrix `especie` where the bacteria in the numerator of the selected principal balance i are placed.
- DemSPBal: List. The component i of the list has the number of the row of the matrix `especie` where the bacteria in the denominator of the selected principal balance i are placed.
- MatrixSPBal: MatrixSPBal is the matrix that contains the covariates of the model. The first line es equal to 1 for all columns. The other rows contain the value of one SPBal at all time points. The selected principal balance of the row i+1 has at its numerator the bacteria placed in the rows `NumSPBal[[i]]` of the "especie". The selected principal balance of the row i+1 has at its denominator the bacteria placed in the rows `DemSPBal[[i]]` of the "especie".
- PercenVarianceSPBal: Vector. The component of the vector i contains the percentage of variance of the SPBal with numerator `NumSPBal[[i]]` and denominator `DemSPBal[[i]]`.

References

Creus-Martí, I., Moya, A., Santonja, F. J. (2022). Bayesian hierarchical compositional models for analysing longitudinal abundance data from microbiome studies. Complexity, 2022.

Examples

```
set.seed(314)
esp2=t(gtools::rdirichlet(n=6, c(1,1,5,1,1)))

Num2<-list(3,c(3,5),1,c(3,5,4))
Dem2<-list(5,4,2,c(1,2))

ObtainigValueSPBal(Num2,Dem2,esp2,6)
```

`ObtainingDIC` *Writting the loglikelihood of the dirichlet*

Description

This function calculates the loglikelihood of the dirichlet for the BPBM model. Then, it calculates the loglikelihood with the parameters of each iteration of the MCMC chains and introduces the values in a vector called vectorD. DIC=(1/2)*var(vectorD)+mean(VectorD)

Usage

```
ObtainingDIC(cadenas, MatrizPBmodelo, E, Tt, especiemodi)
```

Arguments

`cadenas`	Matrix with the iterations (in rows) of all the Markov Chains obtained in th estimation. It is the output of "StudyingParam" adding "$AllChainsJoined".
`MatrizPBmodelo`	Matrix with the covariates of the model. In an example with two SPBal and three time points, the covariates are written in the following order: $\begin{matrix} 1 & 1 & 1 \\ SPBal_{1,t-1} & SPBal_{1,t-2} & SPBal_{1,t-3} \\ SPBal_{2,t-1} & SPBal_{2,t-2} & SPBal_{2,t-3} \end{matrix}$
`E`	Number f bacteria in the dataset.
`Tt`	Number of time points available
`especiemodi`	Matrix that contains at row i the bacterial taxa of bacteria i at time points t=2,...,`Tt`.

Value

Returns a data.frame with the DIC value (using the rule, pD = var/2).

References

Creus-Martí, I., Moya, A., Santonja, F. J. (2022). Bayesian hierarchical compositional models for analysing longitudinal abundance data from microbiome studies. Complexity, 2022.

Examples

```
set.seed(314)
especie=t(gtools::rdirichlet(n=2, c(1,2,3)))
E=3
Tt=2
MatrizPBmodelo=rbind(c(1,1),c(-0.3,0.4),c(0.3,0.5))
set.seed(314)
est=Estimating_BPBM(especie,
                  Tt,
                  E,
                  MatrizPBmodelo,
                  nn.chain=3,
                  nn.burnin=1000,
                  nn.sample=5000,
                  nn.thin=10)
SumFinal=StudyingParam(est$R2jagsOutput$BUGSoutput$summary   ,est$SamplesAllChains)
cadenas=SumFinal$AllChainsJoined
especiemodi=especie[,-1]

ObtainingDIC(cadenas,MatrizPBmodelo,E,Tt,especiemodi)
```

`PBalance` *Calculating balances*

Description

This function calculates the balance that has at the numerator the bacteria placed at `Num` and has at the denominator the bacteria placed at `Dem`

Usage

```
PBalance(A, Num, Dem, especie)
```

Arguments

`A`	Number. The balance will be calculated for t=1,2,...,`A` time points.
`Num`	vector that contains the position in the matrix `especies` of the families that we position at the numerator of the balance.
`Dem`	vector that contains the position in the matrix `especies` of the families that we position at the denumerator of the balance
`especie`	Matrix that contains at row i the bacterial taxa of bacteria i at all time points.

Value

Returns a vector with the value of the balance in each time point.

Examples

```
especie1=cbind(c(0.5,0.3,0.1,0.1), c(0.1,0.3,0.6,0.1))
Num=c(1,2)
Dem=c(3,4)
A=2
PBalance(A,Num,Dem,especie1)
```

`PBalancePredi` *Calculating balances for a composition*

Description

This function calculates the balance that has at the numerator de bacteria placed at `Num` and has at the denominator the bacteria placed at `Dem`

Usage

```
PBalancePredi(Num, Dem, DatosEsperanzas)
```

Arguments

`Num`	vector that contains the position in the matrix `especies` of the families that we position at the numerator of the balance.
`Dem`	vector that contains the position in the matrix `especies` of the families that we position at the denumerator of the balance
`DatosEsperanzas`	Vector with a coda composition. The bacteria are in the same orden than the matrix `especie`

Value

Returns the value of the balance

Examples

```
Num=c(1,2)
Dem=c(3,4)
DatosEsperanzas=c(0.1,0.3,0.4,0.2)

PBalancePredi(Num,Dem,DatosEsperanzas)
```

`PBmodel` *Obtaining the regression value of the BPBM*

Description

This function calculates the value of the BPBM regression, defined by:

$$\mu_{it} = a_{i0} + a_{i1} \cdot \text{SPBal}_{1,t-1} + \cdots + a_{iM} \cdot \text{SPBal}_{M,t-1}$$

Usage

```
PBmodel(A, MatrizPBmodelo, E, Tt)
```

Arguments

`A`	Matrix that contains all the parameters of the model. The parameters are written in the matrix in the following order (in an example with three bacteria):

a10	a11	a12	...	a1M
a20	a21	a22	...	a2M
a30	a31	a32	...	a3M

`MatrizPBmodelo`	Matrix. Output of "ObtainingValueSPBal" called "MatrixSPBal".
`E`	Number of bacteria in the dataset.
`Tt`	Number of time points.

Value

Returns a matrix. The row i contains the regression values of the bacteria i at all time points.

References

Creus-Martí, I., Moya, A., Santonja, F. J. (2022). Bayesian hierarchical compositional models for analysing longitudinal abundance data from microbiome studies. Complexity, 2022.

Examples

```
A=rbind(c(1,2,3),c(4,5,6),c(7,8,9))
MatrizPBmodelo=cbind(c(1,2,3),c(4,5,6),c(7,8,9))
E=3
Tt=3


PBmodel(A,MatrizPBmodelo, E,Tt)
```

`PCAbiplot` *PCA of the estimated parameters*

Description

This function applys a PCA to the estimate parameters (using function "prcomp" with `center = TRUE` and `scale. = TRUE`). Then uses the ggbiplot function to plot the biplot.

Usage

```
PCAbiplot(paramEstimadosFinal, names, E)
```

Arguments

`paramEstimadosFinal`	The estimate parameters, in the following order: a11,a12,a13, a21, a22,a23, ...a(D-1)1,a(D-1)2,a(D-1)3,tau. Where D is the number of bacterial especies present in the matrix `especie`.
`names`	Vector with the name of the bacteria. The component i has the name of the bacteria i, with i=1,...,D. The bacteria in the las position of the vector is the bacteria used as reference in the alr transformation.
`E`	Number of bacteria available.

Value

Returns a list with the PCA biplot, the variance explained of each Principal Component and an object of class "prcomp" with the PCA. In the biplot, "a" denotes the intercept, "b" denotes the parameter that give information about the importance of the bacteria in defining herself in the next time point and "c" denotes denotes the parameter that give information about the importance of the rest of the community in defining the bacteria in the next time point.

Examples

```
set.seed(123)
especie=t(gtools::rdirichlet(10,c(1,3,1,2,4)))
names=c("Bact1","Bact2","Bact3","Bact4","Bact5")
tau1=0.4
parms1= cbind(c(0.1,0.2,0.4,0.6),c(-0.2,0.1,0.1,0.3),c(0.3,0.2,0.3,0.5))
paramEstimadosFinal=c(as.vector( t(parms1)),tau1)

PCAbiplot(paramEstimadosFinal,names,5)
```

`Percen_Variance` *Percentage of variance*

Description

This function calculates the variance of each row of the matrix `PB`. Returns the percentage of variance of each row of the matrix `PB`.

Usage

```
Percen_Variance(PB)
```

Arguments

`PB`	Matrix.

Value

Returns a vector with percentage of variance of each row of the matrix `PB`.

Examples

```
matt=matrix(c(1:4),2,2)


Percen_Variance(matt)
```

`PlotDendogram` *Plotting a dendogram*

Description

Plots the dendogram obtained using the Ward's method for obtaining the principal balances. The process follow in this function is explained in Section 3.1 of (Creus-Martí et al, 2022)

Usage

```
PlotDendogram(especie, names)
```

Arguments

`especie`	Matrix that contains at row i the bacterial taxa of bacteria i at all time points.
`names`	Vector with the names of the bacteria in the same order that are written in `especie`.

Value

Returns a list with: the dendogram.

- Num: List. The component i of the list has the number of the row of the matrix `especie` where the bacteria in the numerator of the principal balance i are placed.
- Dem: List. The component i of the list has the number of the row of the matrix `especie` where the bacteria in the denominator of the principal balance i are placed.
- dendogram: Plots the dendogram.

References

Creus-Martí, I., Moya, A., Santonja, F. J. (2022). Bayesian hierarchical compositional models for analysing longitudinal abundance data from microbiome studies. Complexity, 2022.

Examples

```
names2=c("Bact1","Bact2","Bact3","Bact4","Bact5")
set.seed(314)
esp2=t(gtools::rdirichlet(n=6, c(1,1,5,1,1)))


PlotDendogram(esp2,names2)
```

`PredictionBPBM` *Predicting using BPBM*

Description

This function calculates the expected value and variance of the bacteria at time point `Tt`. Then, this function calculates the expected value and variance of the bacteria at time point t=(`Tt`+1),...,K. It calculates the expected value at each time point for each markov chain iteration. The expected value for each time point is the mean of the expected values of all iterations.mAnalogous with the variance, the dirichlet parameters and the expected valur plus(and minus) two times the sqrt of the variance.

Usage

```
PredictionBPBM(
  NumSPBal,
  DemSPBal,
  MCMC.CHAINS,
  alpha,
  K,
  esperanza,
  Var,
  E,
  Tt,
  MatrizPBmodelo
)
```

Arguments

`NumSPBal`	List. The component i of the list has the number of the row of the matrix `especie` where the bacteria in the numerator of the principal balance i are placed.
`DemSPBal`	List. The component i of the list has the number of the row of the matrix `especie` where the bacteria in the denominator of the principal balance i are placed.
`MCMC.CHAINS`	Matrix with the iterations of all the chains for all the parameters. Each column has all the iteration of one parameter. If the cero is in the center of the credible interval of one parameter all its iteration in the Marchov Chain have the value 0. It is output of the "StudyingParam" function adding "$AllChainsJoined".
`alpha`	Matrix that contains at the row i the Dirichlet parameter of the bacteria i at t=1,2,3,...,`Tt`.
`K`	Number. The function will calculate the value of the expected value and the variance at `Tt` and predict for the time points t=`Tt`+1,..,K. To predict all the time points available at the data we K=dim(especie.All)-1
`esperanza`	Matrix that contains at row i the expected value of the bacterial taxa of bacteria i at t=1,2,3,...,`Tt`-1.

`Var`	Matrix that contains at row i the variance of the bacterial taxa of bacteria i at t=1,2,3,...,`Tt`-1.
`E`	Number of bacteria available
`Tt`	Number of bacteria available
`MatrizPBmodelo`	is the matrix that contains the covariates of the model. The first line es equal to 1 for all columns. The other rows contain the value of one SPBal at all time points. The selected principal balance of the row i+1 has at its numerator the bacteria placed in the rows `NumSPBal[[i]]` of the "especie". The selected principal balance of the row i+1 has at its denominator the bacteria placed in the rows `DemSPBal[[i]]` of the "especie".

Value

Returns a list with:

- ExpectedValue.All: Matrix. Matrix that contains at row i the expected value of the bacteria i at all time points t=1,2,...,K. The bacterias are placed at the same order than in `especies`.
- VarianceValue.All: Matrix. Matrix that contains at row i the variance of the bacteria i at all time points t=1,2,...,K. The bacterias are placed at the same order than in `especies`.
- DirichlerParam.All: Matrix. Matrix that contains at row i the dirichlet parameter of the bacteria i at all time points t=1,2,...,K. The bacterias are placed at the same order than in `especies`.
- ExpVarmas: Matrix. Matrix that contains at row i the expected value plus two times the sqrt(variance) of the bacteria i at all time points t=`Tt`,...,K, the rest of the time points has 0 values. The bacterias are placed at the same order than in `especies`.
- ExpVarmenos: Matrix. Matrix that contains at row i the expected value plus two times the sqrt(variance) of the bacteria i at all time points t=`Tt`,...,K,the rest of the time points has 0 values. The bacterias are placed at the same order than in `especies`.

References

Creus-Martí, I., Moya, A., Santonja, F. J. (2022). Bayesian hierarchical compositional models for analysing longitudinal abundance data from microbiome studies. Complexity, 2022.

Examples

```
NumSPBal=list(1,c(1,2))
DemSPBal=list(2,3)
MCMC.CHAINS=cbind(c(0.1,0.11),
                  c(0.2,0.21),
                  c(0.3,0.31),
                  c(-0.1,-0.11),
                  c(0.15,0.105),
                  c(0.44,0.41),
                  c(0.3,0.31),
                  c(0.201,0.221),
                  c(0.13,0.113) )
alpha=cbind(c(0.1,0.2,0.1),c(0.1,0.5,0.3))
K=3
```

```
esperanza=cbind(c(0.2,0.2,0.6))
Var=cbind(c(0.1,0.01,0.11))
E=3
Tt=2
MatrizPBmodelo=cbind(c(1,0.3,0.2))

PredictionBPBM(NumSPBal,DemSPBal,MCMC.CHAINS, alpha,K,esperanza,Var,E,Tt,MatrizPBmodelo )
```

`PredictionEstParmFunc` *Predicting using dirichl-gLV*

Description

This function calculates the expected value and variance of the bacteria at time point `Tt`. Then, this function calculates the expected value and variance of the bacteria at time point t=(`Tt`+1),...,K

Usage

```
PredictionEstParmFunc(
  paramEstimadosFinal,
  EspecieMaxima,
  alpha,
  K,
  esperanza,
  Var,
  E,
  Tt
)
```

Arguments

`paramEstimadosFinal`

The estimate parameters. Vector equal to `c(tau,as.vector( pam))` where:

- pam Matrix. Each row has the parameters of each bacteria. Following our example, pam has the parameters placed as follows:

$$\begin{array}{ccc} r1 & a11 & a12 \\ r2 & a21 & a22 \end{array}$$

- tau Number. Value of the tau parameter in the model

`EspecieMaxima`	Row in which the bacteria chosen as reference is in `especie`.This bacteria is used as reference in the alr transformation that the model does.
`alpha`	Matrix that contains at the row i the dirichlet parameter of the bacteria i at t=1,2,3,...,`Tt`.

`K`	Number. The function will calculate the value of the expected value and the variance at `Tt` and predict for the time points t=`Tt`+1,..,K. To predict all the time points available at the data we K=dim(especie.All)-1
`esperanza`	Matrix that contains at row i the expected value of the bacterial taxa of bacteria i at t=1,2,3,...,`Tt`-1.
`Var`	Matrix that contains at row i the variance of the bacterial taxa of bacteria i at t=1,2,3,...,`Tt`-1.
`E`	Number of bacteria available
`Tt`	Number of time points available

Details

The regression of this model, in an example with three bacteria, is defined by

$$r_1 \cdot log(x_1(t)/x_3(t)) + log(x_1(t)/x_3(t)) \cdot [a_{11} \cdot log(x_1(t)/x_3(t))(t) + a_{12} \cdot log(x_2(t)/x_3(t))]$$

$$r_2 \cdot log(x_2(t)/x_3(t)) + log(x_2(t)/x_3(t)) \cdot [a_{21} \cdot log(x_1(t)/x_3(t))(t) + a_{22} \cdot log(x_2(t)/x_3(t))]$$

Value

Returns a list with:

- ExpectedValue.All: Matrix. Matrix that contains at row i the expected value of the bacteria i at all time points t=2,...,K. The bacteria are placed at the same order than in `especies`.
- VarianceValue.All: Matrix. Matrix that contains at row i the variance of the bacteria i at all time points t=2,...,K. The bacteria are placed at the same order than in `especies`.
- DirichlerParam.All: Matrix. Matrix that contains at row i the dirichlet parameter of the bacteria i at all time points t=2,...,K. The bacteria are placed at the same order than in `especies`.

References

Creus-Martí, I. and Moya, A. and Santonja, F. J. (2018). A Statistical Model with a Lotka-Volterra Structure for Microbiota Data. Lucas Jodar, Juan Carlos Cortes and Luis Acedo, Modelling or engineering and human behavior 2018, Instituto Universitario de Matematica Multidisciplinar. ISBN: 978-84-09-07541-6

Examples

```
pam.ini=rbind(c(0.1,0.2,0.3),c(0.4,0.5,0.6))
paramEstimadosFinal=c(5, as.vector(pam.ini))
EspecieMaxima=3
alpha=cbind(c(2,2,3),c(1,1,3))
K=3
esperanza=cbind(c(0.2,0.3,0.5))
Var=cbind(c(0.2,0.3,0.5))
E=3
Tt=2
```

```
PredictionEstParmFunc(paramEstimadosFinal,EspecieMaxima, alpha,K,esperanza,Var,E,Tt )
```

`PredictionFBM` *Predicting using FBM*

Description

This function calculates the expected value and variance of the bacteria at time point `Tt`. Then, this function calculates the expected value and variance of the bacteria at time point t=(`Tt`+1),...,K

Usage

```
PredictionFBM(
  paramEstimadosFinal,
  EspecieMaxima,
  alpha,
  K,
  esperanza,
  Var,
  E,
  Tt
)
```

Arguments

`paramEstimadosFinal`	The estimate parameters, in the following order: a11,a12,a13, a21, a22,a23, ...a(D-1)1,a(D-1)2,a(D-1)3,tau. Where D is the number of bacterial species present in the matrix `especie`.
`EspecieMaxima`	Row in which the bacteria chosen as reference is in `especie`.This bacteria is used as reference in the alr transformation that the model does and it is placed at the denominator of the balance)
`alpha`	Matrix that contains at the row i the Dirichlet parameter of the bacteria i at t=1,2,3,...,`Tt`.
`K`	Number. The function will calculate the value of the expected value and the variance at `Tt` and predict for the time points t=`Tt`+1,..,K. To predict all the time points available at the data we K=dim(especie.All)-1
`esperanza`	Matrix that contains at row i the expected value of the bacterial taxa of bacteria i at t=1,2,3,...,`Tt`-1.
`Var`	Matrix that contains at row i the variance of the bacterial taxa of bacteria i at t=1,2,3,...,`Tt`-1.
`E`	Number of bacteria available
`Tt`	Number of bacteria available

Details

The regression of this model is defined by

$\mu_{it} = a_{i1} + a_{i2} \cdot \text{alr}(x_{i,(t-1)}) + a_{i3} \cdot \text{Balance}(x_{i,(t-1)})$ for $i = 1, \dots, D-1$ where D is the number of bacteria

Value

Returns a list with:

- ExpectedValue.All: Matrix. Matrix that contains at row i the expected value of the bacteria i at all time points t=1,2,...,K. The bacteria are placed at the same order than in `especies`.
- VarianceValue.All: Matrix. Matrix that contains at row i the variance of the bacteria i at all time points t=1,2,...,K. The bacteria are placed at the same order than in `especies`.
- DirichlerParam.All: Matrix. Matrix that contains at row i the dirichlet parameter of the bacteria i at all time points t=1,2,...,K. The bacteria are placed at the same order than in `especies`.

References

Creus-Martí, I., Moya, A., Santonja, F. J. (2021). A Dirichlet autoregressive model for the analysis of microbiota time-series data. Complexity, 2021, 1-16.

Examples

```
Tt=2
E=3
tau=5
EspecieMaxima=3
K=3
parms11=c(0.1,0.2,0.3,0.4,0.5,0.6,tau)
alpha=cbind(c(1.726793,1.892901,1.380306),
          c(1,1,3))
Expected=cbind(c(alpha[1,1]/tau, alpha[2,1]/tau, alpha[3,1]/tau  ),
             c(alpha[1,2]/tau,alpha[2,2]/tau,alpha[3,2]/tau))
Variance=cbind(c(0.03768101, 0.03920954, 0.03330857 ),
             c( 0.03683242,0.02784883, 0.0413761 ))
Expected.final=Expected[,-2]
Variance.final=Variance[,-2]

PredictionFBM(parms11,EspecieMaxima, alpha,K,Expected.final,Variance.final,E,Tt )
```

`PreparingTheData` *Preparing dataset*

Description

Preparing the dataset to be introduce in the models' functions. In order to introduce the usage of the package there is a README file. You can find the link to the file using `base::system.file("extdata", "README.pdf", package = "CoDaLoMic")`. On windows you can open the file with `base::shell.exec(system.file("ext` `"README.pdf", package = "CoDaLoMic"))`.

Usage

```
PreparingTheData(DaTa, Pred)
```

Arguments

`DaTa`	data.frame. The first column contains the time point information (natural numbers 1,2,3...). The rest of the columns contain the relative abundance of each bacteria at the different time points. The values of each column must sum 1.
`Pred`	Number. The data at t=1,...,Pred-1 will be used to estimate the model. The rest of the time points will be used to study the capacity of the model to predict. If `Pred==0` all the datatset will be used to estimate the model.

Value

If `Pred==0` returns a list with

- `Tt` - The number of time points available.
- `E` - Number of bacteria available
- `especieOriginal` - Matrix that contains at row i the bacterial taxa of bacteria i at all time points.
- `especiemodiOriginal` - Matrix that contains at row i the bacterial taxa of bacteria i at time points t=2,...,`Tt`.

If `Pred!=0` returns a list with

- `Tt` - The number of time points available used to estimate the model (`Tt`=Pred-1).
- `E` - Number of bacteria available
- `especieOriginal` - Matrix that contains at row i the bacterial taxa of bacteria i at the time points t=1,2,...,Pred-1.
- `especiemodiOriginal` - Matrix that contains at row i the bacterial taxa of bacteria i at time points t=2,...,Pred-1.
- `especieOriginal.All` - Matrix that contains at row i the bacterial taxa of bacteria i at the time points.
- `especiemodiOriginal.All` - Matrix that contains at row i the bacterial taxa of bacteria i at time points.
- `K` - Number of time points available at the dataset.

Examples

```
df<-data.frame(cbind(c(1,2,3),
                     c(0.5,0.2,0.3),
                     c(0.2,0.1,0.6),
                     c(0.1,0.1,0.8),
                     c(0.3,0.3,0.4)))
PreparingTheData(df,Pred=0)

df2<-data.frame(cbind(c(1,2,3,4,5),
                      c(0.1,0.1,0.1,0.2,0.5),
                      c(0.2,0.2,0.2,0.2,0.2),
                      c(0.2,0.3,0.1,0.2,0.2)))
PreparingTheData(df2,Pred=4)
```

`QualityControl` *Analysing the quality of the estimation*

Description

This function calculates the root-mean-square deviation (RMSD), the Nash Sutchiffe Coefficient, the residual sum of squares (RSS) and the mean absolute percentage error (MAPE) for the matrices introduces. This function also calculates the mean of the RMSD, the mean of the Nash Sutchiffe Coefficient and the mean of the RSS.

Usage

```
QualityControl(matrixData, matrixExpected, names.especie)
```

Arguments

`matrixData`	Matrix that contains at row i the bacterial taxa of bacteria i at the time points that we want take into account to calculate the quality control values.
`matrixExpected`	Matrix that contains at row i the expected value of the bacterial taxa i at the time points that we want take into account to calculate the quality control values. The bacteria must be placed in the same order than in `matrixData`
`names.especie`	Vector with the names of the bacteria in the same order that are placed in the `matrixData` matrix.

Value

Returns a data.frame.

Examples

```
names.especie=c("Bact1", "Bact2", "Bact3")
matrixExpected=matrix(c(1:9),3,3)
matrixData=matrixExpected+0.1

QualityControl(matrixData, matrixExpected,names.especie)
```

`ridgeregression` *Ridge regression*

Description

Ridge regression

Usage

```
ridgeregression(Tt, especie, E, EspecieMaxima, seed = NULL)
```

Arguments

`Tt`	Number of time points available
`especie`	Matrix that contains at row i the bacterial taxa of bacteria i at all time points. The bacteria placed in the last row of the matrix will be used as reference in the alr transformation and will be at the denominator of the balance.
`E`	Number of bacteria available.
`EspecieMaxima`	Row in which the bacteria used as reference is in `especie`. This is the bacteria that is going to be at the denominator of the balance and the denominator of the alr transformation. As a result, in this function, `EspecieMaxima` must be equal to `E`
`seed`	Number. Set a seed. Default `seed=NULL`.

Value

Returns the result of the ridge regression, object of class "ridgelm".

Examples

```
set.seed(123)
especie=t(gtools::rdirichlet(10,c(1,3,1,2,4)))
Tt=10
E=5
EspecieMaxima=5
```

```
ridgeregression(Tt,especie, E, EspecieMaxima, 558562316)
```

`rxnrate` *Solving the right side of the gLV equations*

Description

This function calculates the right side of the gLV equation.

Usage

```
rxnrate(State, parms)
```

Arguments

`State`	Vector with a CoDa composition
`parms`	Matrix. Each row has the parameters of each differential equation. following our example, parms has the parameters placed as follows:

r1	a11	a12	a13
r2	a21	a22	a23
r3	a31	a32	a33

Details

For instance, if we want to solve the following gLV equations:

$$\frac{dx_1(t)}{dt} = r_1 \cdot x_1(t) + x_1(t) \cdot [a_{11} \cdot x_1(t) + a_{12} \cdot x_2(t) + a_{13} \cdot x_3(t)]$$

$$\frac{dx_2(t)}{dt} = r_2 \cdot x_2(t) + x_2(t) \cdot [a_{21} \cdot x_1(t) + a_{22} \cdot x_2(t) + a_{23} \cdot x_3(t)]$$

$$\frac{dx_3(t)}{dt} = r_3 \cdot x_3(t) + x_3(t) \cdot [a_{31} \cdot x_1(t) + a_{32} \cdot x_2(t) + a_{33} \cdot x_3(t)]$$

This function returns a vector with the value of:

$$r_1 \cdot x_1(t) + x_1(t) \cdot [a_{11} \cdot x_1(t) + a_{12} \cdot x_2(t) + a_{13} \cdot x_3(t)]$$

$$r_2 \cdot x_2(t) + x_2(t) \cdot [a_{21} \cdot x_1(t) + a_{22} \cdot x_2(t) + a_{23} \cdot x_3(t)]$$

$$r_3 \cdot x_3(t) + x_3(t) \cdot [a_{31} \cdot x_1(t) + a_{32} \cdot x_2(t) + a_{33} \cdot x_3(t)]$$

Value

Returns a vector with the value of the right side of the gLV equations.

Examples

```
cinit1<-c(x1<-0.7,x2<-0.2,x3<-0.1)
parms1= cbind(c(0.1,0.2,-0.1),c(-0.2,0.1,-0.1),c(0.3,0.2,0.3),c(0.1,0.22,0.2))
rxnrate(cinit1,parms1)
```

`Simulated` *Gut microbiome simulated dataset*

Description

Simulated dataset with 5 microbial taxa and 10 time points. Following the scheme given by Faust et al (2018), we generated the interaction matrix using the algorithm proposed by Klemm and Eguíluz (2002), and we generated the initial abundances using the Poisson distribution. With these two tools, we simulated the data using the generalized Lotka-Volterra structure. We carried out the simulation using the R package seqtime (Faust et al, 2018). Focusing on technical details, to generate the interaction matrix we set the clique size at 4, the diagonal values at -1, the interaction connectance at 0.04, the positive edge percentage at 64

Usage

```
data(Simulated)
```

Format

A data frame with 10 rows and 6 columns.

References

- K. Faust, F. Bauchinger, B. Laroche et al., “Signatures ofecological processes in microbial community time series”. Microbiome, vol. 6, no. 1, p. 120, 2018
- K. Klemm and V. M. Eguíluz, “Growing scale-free networkswith small-world behavior”. Physical Review, vol. 65, no. 5,Article ID 057102, 2002.

`StudyingParam` *Controlling quality of the convergence in BPBM*

Description

This function controls that the value of the Rhat is between 0.9 and 1.1. In addition, it controls that the effective sample size is bigger than 100 and that the zero is not at the center of the credible interval (the interval between 2.5 and 97.5). We consider that the zero is in the center of the credible interval when the zero is between the 25 and the 75 quantile of the distribution formed by the limits if the credible interval.

Usage

```
StudyingParam(Sum, MCMC.CHAINS)
```

Arguments

`Sum`	Matrix with the summary of the "Estimating_BPBM".It is the output of the "Estimating_BPBM" adding "$R2jagsOutput$BUGSoutput$summary".
`MCMC.CHAINS`	Matrix with the values of all the Markov chains for all parameters. It is the output of the "Estimating_BPBM" adding "$SamplesAllChains".

Value

Returns a list with:

- Param.Summary: Matrix. The matrix Sum with a zero in the column `Sum[,"mean"]` when a parameter has the zero in the center of its credible interval.
- AllChainsJoined: Matrix. The matrix MCMC.CHAINS with a zero in all the iterations of the chain when a parameter has the zero in the center of its credible interval.

Examples

```
set.seed(314)
especie=t(gtools::rdirichlet(n=6, c(6,6,1,6,6)))
E=5
Tt=6
MatrizPBmodelo=rbind(c(1,1,1,1,1,1),c(-0.3,0.4,0.3,-0.7,-0.4,-0.6),c(0.3,0.5,-0.3,0.1,0.4,0.1))

est=Estimating_BPBM(especie,
                   Tt,
                   E,
                   MatrizPBmodelo,
                   nn.chain=3,
                   nn.burnin=1000,
                   nn.sample=5000,
                   nn.thin=10)

StudyingParam(est$R2jagsOutput$BUGSoutput$summary,est$SamplesAllChains)
```

`TableBPBM` *Obtaining a table with the SPBal information*

Description

Returns a table with the percentage of variance that each SPBal has, the bacteria that goes in the numerator and denominator of the balance, the relationship between the group in the numerator and the denominator and the bacteria most influences by this SPBal.

Usage

```
TableBPBM(
  NumSPBal,
  DemSPBal,
  PerVar,
  MatrizPBmodelo,
  Estimated.Param,
  BB = 0.55,
  names,
  E
)
```

Arguments

`NumSPBal`	List. Output of "ObtainigValueSPBal" function.List. The component i of the list has the number of the row of the matrix `especie` where the bacteria in the numerator of the selected principal balance i are placed.
`DemSPBal`	List. Output of "ObtainigValueSPBal" function.List. The component i of the list has the number of the row of the matrix `especie` where the bacteria in the denominator of the selected principal balance i are placed.
`PerVar`	Vector. Output of "ObtainigValueSPBal" function. The component of the vector i contains the percentage of variance of the SPBal with numerator `NumSPBal[[i]]` and denominator `DemSPBal[[i]]`.
`MatrizPBmodelo`	Matrix. Output of "ObtainigValueSPBal" function. MatrixSPBal is the matrix that contains the covariates of the model. The first line es equal to 1 for all columns. The other rows contain the value of one SPBal at all time points. The selected principal balance of the row i+1 has at its numerator the bacteria placed in the rows `NumSPBal[[i]]` of the "especie". The selected principal balance of the row i+1 has at its denominator the bacteria placed in the rows `DemSPBal[[i]]` of the "especie".
`Estimated.Param`	Vector. Column "mean" of the output of StudyingParam function.
`BB`	The bacteria in the numerator and the denominator of the balance are considered similar if the mean of the SPBal is between (-BB,BB).Default: 0.55.
`names`	Vector with the bacteria's names placed in the order in which appear in the rows of the matrices `especie` and `especiemodi`.
`E`	number of bacteria in the dataset.

Value

Returns the table as formatted text. By using the `cat` function, you can generate a LaTeX-ready table that can be easily copied and pasted.

Examples

```
NumSPBal=list(c(3,4),3,2)
DemSPBal=list(1,5,4)
```

```
PerVar=c(41.37487, 21.08270, 19.16870)
MatrizPBmodelo=rbind(c(1.00000000,  1.0000000,  1.0000000,  1.0000000 , 1.0000000),
                   c(1.84449081, -1.3569851, -0.1388348, -0.5269079, -1.3288684),
                   c(0.27531685,  0.4741394, -1.9045554, -0.3581268, -0.4768543),
                   c( 0.07782991,  0.3492473, -0.7138882,  1.4332828, -0.7203295))
namesOr=c("Bact1", "Bact2","Bact3","Bact4","Bact5")
Estimated.Param=c(0.5, 0.3, -0.2,0.1)
E=5

tat=TableBPBM(NumSPBal,DemSPBal,PerVar,MatrizPBmodelo,Estimated.Param,BB=0.55,namesOr, E)
cat(tat, sep = "\n")
```

`TableFBM` *Obtainig a table with the interpretable parameters*

Description

This function returns a table with the interpretable parameters of the FBM model.

Usage

```
TableFBM(paramEstimadosFinal, names, E)
```

Arguments

`paramEstimadosFinal`	The estimate parameters, in the following order: a11,a12,a13, a21, a22,a23, ...a(D-1)1,a(D-1)2,a(D-1)3,tau. Where D is the number of bacterial species present in the matrix `especie`.
`names`	Vector of length D. The component i has the name of the bacteria i.
`E`	Number of bacteria available.

Details

$\mu_{it} = a_{i1} + a_{i2} \cdot \text{alr}(x_{i,(t-1)}) + a_{i3} \cdot \text{Balance}(x_{i,(t-1)})$ for $i = 1, \dots, D-1$ where D is the number of bacteria

Value

Returns the table as formatted text. By using the `cat` function, you can generate a LaTeX-ready table that can be easily copied and pasted.

References

Creus-Martí, I., Moya, A., Santonja, F. J. (2021). A Dirichlet autoregressive model for the analysis of microbiota time-series data. Complexity, 2021, 1-16.

Examples

```
paramEstimadosFinal=c(1,2,3,1,2,3,1,2,3)
names=c("Bact1", "Bact2","Bact3")
E=3

tat=TableFBM(paramEstimadosFinal,names,E)
cat(tat, sep = "\n")
```

`Table_alr_Dirich_glv` *Obtainig a table with the interpretable parameters*

Description

This function returns a table with the interpretable parameters of the Dirich-gLV model.

Usage

```
Table_alr_Dirich_glv(Param.Estimates, especie, names, E)
```

Arguments

`Param.Estimates`	Vector with the estimates parameters. It is equal to `c(tau,as.vector( pam))` where: • pam Matrix. Each row has the parameters of each bacteria. Following our example, pam has the parameters placed as follows: r1 a11 a12 r2 a21 a22 • tau Number. Value of the tau parameter in the model
`especie`	Matrix that contains at row i the bacterial taxa of bacteria i at all time points. The bacteria placed in the last row of this matrix is the one used as reference in the alr transformation that the model applies.
`names`	Vector with the name of the bacteria in the same order than are present in the `especie` matrix.
`E`	Number of bacteria available.

Details

In an example with three bacteria, the regression of this model is defined by

$$r_1 \cdot log(x_1(t)/x_3(t)) + log(x_1(t)/x_3(t)) \cdot [a_{11} \cdot log(x_1(t)/x_3(t))(t) + a_{12} \cdot log(x_2(t)/x_3(t))]$$

$$r_2 \cdot log(x_2(t)/x_3(t)) + log(x_2(t)/x_3(t)) \cdot [a_{21} \cdot log(x_1(t)/x_3(t))(t) + a_{22} \cdot log(x_2(t)/x_3(t))]$$

Value

Returns the table as formatted text. By using the `cat` function, you can generate a LaTeX-ready table that can be easily copied and pasted.

References

Creus-Martí, I. and Moya, A. and Santonja, F. J. (2018). A Statistical Model with a Lotka-Volterra Structure for Microbiota Data. Lucas Jodar, Juan Carlos Cortes and Luis Acedo, Modelling or engineering and human behavior 2018, Instituto Universitario de Matematica Multidisciplinar. ISBN: 978-84-09-07541-6

Examples

```
pam.ini=rbind(c(0.1,0.2,0.3),c(0.4,0.5,0.6))
paramEstimadosFinal=c(5, as.vector(pam.ini))
E=3
especie=cbind(c(0.2,0.4,0.4),c(0.1,0.1,0.8),c(0.5,0.1,0.4))
names=c("a","b","c")

tat=Table_alr_Dirich_glv(paramEstimadosFinal,especie,names,E)
cat(tat, sep = "\n")
```

`TauAndParameters_EstParmFunc_FBM` *Obtaining the value of tau and the estimate value of the rest of the parameters*

Description

This function estimates the parameters of the FBM model.

Usage

```
TauAndParameters_EstParmFunc_FBM(
  ttau = 30,
  ridge.final,
  Iter.EstParmFunc = 80,
  especie,
  EspecieMaxima,
  Tt,
  E,
  seed = NULL
)
```

Arguments

`ttau`	Number. We estimate de FBM model for the values of tau: 1, 2,..., ttau
`ridge.final`	Object of class "ridgelm". Values obtained with the ridge regression.
`Iter.EstParmFunc`	Number. Number of iterations. Default: 80 iterations.
`especie`	Matrix that contains at row i the bacterial taxa of bacteria i at all time points. The bacteria placed in the last row of the matrix will be used as reference in the alr transformation and will be at the denominator of the balance.
`EspecieMaxima`	Row in which the bacteria used as reference is in `especie`. This is the bacteria that is going to be at the denominator of the balance and at the denominator of the alr transformartion. As a result, in this function, `EspecieMaxima` must be equal to `E`
`Tt`	Number of time points available
`E`	Number. Number of bacteria available.
`seed`	Number. Set a seed. Default `seed=NULL`.

Details

We give to the parameter tau the value 1,2,...,`ttau`. We estimate the FBM model for all this values (using the function "Estimate_param_FBM") and we select the value of tau that minimizes the AIC. The regression of this model is defined by

$$\mu_{it} = a_{i1}+a_{i2}\cdot\text{alr}(x_{i,(t-1)})+a_{i3}\cdot\text{Balance}(x_{i,(t-1)}) \text{ for } i = 1,\dots,D-1 \text{ where } D \text{ is the number of bacteria}$$

Value

Returns a list with:

- EstimateParameters: Vector with the estimated parameters, in the following order: a11,a12,a13, a21, a22,a23, ...a(D-1)1,a(D-1)2,a(D-1)3,tau. Where D is the number of bacterial species present in the matrix `especie`.
- AIC Number: Value of the AIC.
- All.iter: Matrix. Each row has the parameters obtained in each iteration. The parameters are in the columns written in the same order that they are written in `Param.Estimates` . In this matrix we must observe that in the last iterations the values has really similar or equal values, if not, we need to increase the value of `Iter.EstParmFunc`.

References

Creus-Martí, I., Moya, A., Santonja, F. J. (2021). A Dirichlet autoregressive model for the analysis of microbiota time-series data. Complexity, 2021, 1-16.

Examples

```
set.seed(123)
especie=t(gtools::rdirichlet(5,c(1,3,1)))
Tt=5
E=3
EspecieMaxima=3
ridge.final=ridgeregression(Tt,especie, E, EspecieMaxima)
ttau=10
Iter.EstParmFunc=10

TauAndParameters_EstParmFunc_FBM(ttau,ridge.final,Iter.EstParmFunc, especie,EspecieMaxima,Tt,E,714)
```

`vecttor` *Alr of a bacteria*

Description

Writtes a vector with the alr transformation of the bacteria `i` at time points t=2,...,`Tt`.

Usage

```
vecttor(i, especie, Tt, EspecieMaxima)
```

Arguments

`i`	Number. Position of the bacteria that we make the alr in the matrix `especie`. `i` must be different that `EspecieMaxima`.
`especie`	Matrix that contains at row i the bacterial taxa of bacteria i at all time points. The bacteria placed in the last row of the matrix will be used as reference in the alr transformation and will be at the denominator of the balance.
`Tt`	Number of time points available
`EspecieMaxima`	Row in which the bacteria used as reference is in `especie`. This is the bacteria that is going to be at the denominator of the balance and the denominator of the alr transformation. As a result, in this function, `EspecieMaxima` must be equal to `E`

Value

Returns a vector with the alr transformation of the bacteria `i` at time points t=2,...,`Tt`.

Examples

```
set.seed(123)
especie=t(gtools::rdirichlet(10,c(1,3,1,2,4)))
Tt=10
EspecieMaxima=5
```

```
i=2

vecttor(i,especie,Tt,EspecieMaxima)
```

`ZeroData` *Zero replacement*

Description

In this function the zeros are removed or replaced using functions of "zCompositions" package that can be used with longitudinal data (because they do not use the information of other rows to make the replacement).

Usage

```
ZeroData(DaTa, method = "multKM", seed = NULL)
```

Arguments

`DaTa`	data.frame. The first column contains the time point information (natural numbers 1,2,3...). The rest of the columns contain the relative abundance of each bacteria at the different time points. The values of each column must sum 1.
`method`	Character. • If `method="multKM"` - The replacement is carried out with the "multiplicative Kaplan-Meier smoothing spline replacement" (Palarea-Albaladejo and Martín-Fernandez, 2015). Default method. The zeros must be written with a 0. • If `method="multRepl"` - The replacement is carried out with the "multiplicative simple replacement" (Palarea-Albaladejo and Martín-Fernandez, 2015). The zeros must be written with a 0. • If `method="nozeros"` - The bacteria that contains zeros are removed. One column is added to the dataset called "Other".
`seed`	Number. Set a seed. Default `seed=NULL`.

Value

The dataset without zeros.

References

Palarea-Albaladejo J. and Martín-Fernandez JA. zCompositions – R package for multivariate imputation of left-censored data under a compositional approach. Chemometrics and Intelligent Laboratory Systems 2015; 143: 85-96.

Examples

```
set.seed(2)
dat=gtools::rdirichlet(6,c(1,2,3,1,2,3))
dat2=dat
dat2[2,1]=0
dat2[2,2]=dat[2,1]+dat[2,2]
dat2[4,3]=0
dat2[4,4]=dat[4,3]+dat[4,4]

X <- cbind( c(1:6) ,dat2)

Final=ZeroData(X,"multKM",1)
Final2=ZeroData(X,"multRepl",1)
```

Index

∗ **datasets**
    cockroach, 7
    Simulated, 51

B1MODImatrizP, 4
B1MODImodel, 5
Balance, 6
BPBM_Matrix, 6

cockroach, 7
CoDaLoMic *(*CoDaLoMic-package*)*, 3
CoDaLoMic-package, 3

Estimate_Param_EstParmFunc, 8
Estimate_Param_FBM, 9
Estimating_BPBM, 11
EstParmFunc, 13
EstParmFunc_FBM, 14
ExpectedValue_EstParmFunc, 18
ExpectedValues_EstParmFunc_FBM, 16
ExpectedValuess_BPBM, 15

FromVectorToMatrix_BPBM, 19
FVectorPBmodeloPredi, 20

Graphics, 21
GraphicsPrediction, 22
GraphicsPredictionBPBM, 24
GraphicsSPBal, 27

LogVeroFuncBUENA, 28

MaxBacteria, 29
MaxBacteriaPred, 30

ObtainigValuePB, 32
ObtainigValueSPBal, 33
ObtainingDIC, 34

PBalance, 36
PBalancePredi, 36
PBmodel, 37
PCAbiplot, 38
Percen_Variance, 39
PlotDendogram, 40
PredictionBPBM, 41
PredictionEstParmFunc, 43
PredictionFBM, 45
PreparingTheData, 47

QualityControl, 48

ridgeregression, 49
rxnrate, 50

Simulated, 51
StudyingParam, 51

Table_alr_Dirich_glv, 55
TableBPBM, 52
TableFBM, 54
TauAndParameters_EstParmFunc_FBM, 56

vecttor, 58

ZeroData, 59