Package: MiSPU (via r-universe)

September 30, 2024

Type Package

Title Microbiome Based Sum of Powered Score (MiSPU) Tests

Version 1.0

Date 2016-02-29

Author Chong Wu, Wei Pan

Maintainer Chong Wu <wuxx0845@umn.edu>

Description There is an increasing interest in investigating how the compositions of microbial communities are associated with human health and disease. In this package, we present a novel global testing method called aMiSPU, that is highly adaptive and thus high powered across various scenarios, alleviating the issue with the choice of a phylogenetic distance. Our simulations and real data analysis demonstrated that aMiSPU test was often more powerful than several competing methods while correctly controlling type I error rates.

License GPL-2

Imports Rcpp (>= 0.12.1)

Depends R (>= 3.2.3), vegan, ape, aSPU,cluster

Suggests ade4

LinkingTo Rcpp, RcppArmadillo

NeedsCompilation yes

Repository CRAN

Date/Publication 2016-03-18 00:08:39

Contents

`MiSPU-package` *Microbiome Based Sum of Powered Score (MiSPU) Tests*

Description

There is an increasing interest in investigating how the compositions of microbial communities are associated with human health and disease. In this package, we present a novel global testing method called aMiSPU, that is highly adaptive and thus high powered across various scenarios, alleviating the issue with the choice of a phylogenetic distance. Our simulations and real data analysis demonstrated that aMiSPU test was often more powerful than several competing methods while correctly controlling type I error rates.

Details

Package:	MiSPU
Type:	Package
Version:	1.0
Date:	2016-02-29
License:	GPL-2

Author(s)

Chong Wu, Wei Pan Maintainer: Chong Wu <wuxx0845@umn.edu>

References

Pan, W., et al.(2014) A powerful and adaptive association test for rare variants, Genetics, 197(4), 1081-95

Chong, W., Pan, W. (2015) An Adaptive Association Test for Microbiome Data, submitted.

Examples

```
data(throat.otu.tab)
data(throat.tree)
data(throat.meta)

Y.tmp =throat.meta[,3]
Y = rep(0,dim(throat.meta)[1])
Y[Y.tmp=="Smoker"] = 1

cov.tmp = throat.meta[,c(10,12)]
cov = matrix(1,dim(throat.meta)[1],2)
cov[cov.tmp[,1]== "None",1] = 0
cov[cov.tmp[,2]== "Male",2] = 0

start.time = proc.time()
X = as.matrix(throat.otu.tab)

out = MiSPU(Y,X, throat.tree,cov,model =  "binomial", pow = c(2:8, Inf), n.perm = 1000)
out
```

---

corespLeaves *Finding the correspondence OTUs for each taxa*

Description

Finding the descendants for each taxa.

Usage

```
corespLeaves(tree)
```

Arguments

tree	Rooted phylogenetic tree of R class "phylo"

Value

A list containing the descendants for each taxa.

Author(s)

Chong Wu

References

Chong, W., Pan, W. (2015) An Adaptive Association Test for Microbiome Data, submitted.

Examples

```
data(throat.tree)
correspLeaves(throat.tree)
```

`Dirichlet` *The Dirichlet Distribution*

Description

Density function and random number generation for the Dirichlet distribution

Usage

```
rdirichlet(n, alpha)
```

Arguments

`n`	number of random observations to draw
`alpha`	the Dirichlet distribution's parameters. Can be a vector (one set of parameters for all observations) or a matrix (a different set of parameters for each observation), see "Details"

Details

The Dirichlet distribution is a multidimensional generalization of the Beta distribution where each dimension is governed by an α-parameter. Formally this is

$$\mathcal{D}(\alpha_i) = \left[\Gamma(\sum_i \alpha_i) \Big/ \prod_i \Gamma(\alpha_i) \right] \prod_i y_i^{\alpha_i - 1}$$

Usually, `alpha` is a vector thus the same parameters will be used for all observations. If `alpha` is a matrix, a complete set of α-parameters must be supplied for each observation.

Value

returns a matrix with random numbers according to the supplied alpha vector or matrix.

Author(s)

Chong Wu

Examples

```
X1 <- rdirichlet(100, c(5, 5, 10))
X1
```

DirMultOutput *The estimate of Dirichlet-multinomial distribution*

Description

The estimate of Dirichlet-multinomial distribution. It just intermidates estimate.

Usage

```
data(dd)
```

Format

The format is: matrix

Details

It just intermidates estimate. Not very useful.

Examples

```
data(dd)
```

GUniFrac *Generalized UniFrac distances for comparing microbial communities.*

Description

A generalized version of commonly used UniFrac distances. It is defined as:

$$d^{(\alpha)} = \frac{\sum_{i=1}^{m} b_i (p_i^A + p_i^B)^\alpha \left| \frac{p_i^A - p_i^B}{p_i^A + p_i^B} \right|}{\sum_{i=1}^{m} b_i (p_i^A + p_i^B)^\alpha},$$

where m is the number of branches, b_i is the length of ith branch, p_i^A, p_i^B are the branch proportion for community A and B.

Generalized UniFrac distance contains an extra parameter α controlling the weight on abundant lineages so the distance is not dominated by highly abundant lineages. $\alpha = 0.5$ has overall the best power.

Usage

```
GUniFrac(otu.tab, tree,alpha = c(0,0.5,1))
```

Arguments

otu.tab	OTU count table, row - n sample, column - q OTU
tree	Rooted phylogenetic tree of R class "phylo"
alpha	Parameter controlling weight on abundant lineages

Value

Return a list containing

d0	UniFrac(0)
d5	UniFrac(0.5)
d1	UniFrac(1), weighted UniFrac

or a list containing

GUniFrac	The distance matrix for different alpha
alpha	The weight

Note

The time consuming part is written in C and faster than the original one. The function only accepts rooted tree.

Author(s)

Chong Wu <chongwu@umn.edu>

References

Chen, Jun, et al (2012). "Associating microbiome composition with environmental covariates using generalized UniFrac distances." Bioinformatics 28(16):2106-2113.

Examples

```
data(throat.otu.tab)
data(throat.tree)
data(throat.meta)

groups <- throat.meta$SmokingStatus

# Calculate the UniFracs
unifracs <- GUniFrac(throat.otu.tab, throat.tree)
unifracs
```

MiSPU *microbiome based sum of powered score (MiSPU)*

Description

We propose a class of microbiome based sum of powered score (MiSPU) tests based on a newly defined generalized taxon proportion that combines observed microbial composition information with phylogenetic tree information. Different from the existing methods, a MiSPU test is based on a weighted score of the generalized taxon proportion in a general framework of regression, upweighting more likely to be associated microbial lineages. Our simulations demonstrated that one or more MiSPU tests were more powerful than MiRKAT while correctly controlling type I error rates. An adaptive MiSPU (aMiSPU) test is proposed to combine multiple MiSPU tests with various weights, approximating the most powerful MiSPU for a given scenario, consequently being highly adaptive and high powered across various scenarios.

Usage

```
MiSPU(y, X, tree, cov = NULL,model = c("gaussian", "binomial"),
 pow = c(2:8, Inf), n.perm = 1000)
```

Arguments

y	Outcome of interest. It can be a disease indicator; =0 for controls, =1 for cases. Or it can be a quantitative trait. A vector with length n (number of observations).
X	OTU count table, row - n sample, column - q OTU
tree	Rooted phylogenetic tree of R class "phylo"
cov	Covariates. A matrix with dimension n by p (n :number of observation, p : number of covariates).
model	Use "gaussian" for a quantitative trait, and use "binomial" for a binary trait.
pow	The gamma which controls the weight. Larger pow puts more weight on the variables that have larger absolute score.
n.perm	number of permutations or bootstraps.

Value

A list object, including the results for MiSPU_u, MiSPU_w and aMiSPU.

Author(s)

Chong Wu

References

Pan, W., et al.(2014) A powerful and adaptive association test for rare variants, Genetics, 197(4), 1081-95

Chong, W., Pan, W. (2015) An Adaptive Association Test for Microbiome Data, submitted.

Examples

```
data(throat.otu.tab)
data(throat.tree)
data(throat.meta)

Y.tmp =throat.meta[,3]
Y = rep(0,dim(throat.meta)[1])
Y[Y.tmp=="Smoker"] = 1

cov.tmp = throat.meta[,c(10,12)]
cov = matrix(1,dim(throat.meta)[1],2)
cov[cov.tmp[,1]== "None",1] = 0
cov[cov.tmp[,2]== "Male",2] = 0

start.time = proc.time()
X = as.matrix(throat.otu.tab)

out = MiSPU(Y,X, throat.tree,cov,model =  "binomial", pow = c(2:8, Inf), n.perm = 1000)
out
```

`ranking` *ranking the OTUs*

Description

Ranking the importance of each taxa.

Usage

```
ranking(y, X, tree, cov = NULL,gamma,g.taxon.index,model = "binomial")
```

Arguments

`y`	Outcome of interest. It can be a disease indicator; =0 for controls, =1 for cases. Or it can be a quantitative trait. A vector with length n (number of observations).
`X`	OTU count table, row - n sample, column - q OTU
`tree`	Rooted phylogenetic tree of R class "phylo"
`cov`	Covariates. A matrix with dimension n by p (n :number of observation, p : number of covariates).
`gamma`	The best gamma selected by aMiSPU test.
`g.taxon.index`	g.taxon.index = 1 stands for weigted generalized taxon proportion; otherwise means unweighted generalized taxon proportion.
`model`	Use "gaussian" for a quantitative trait, and use "binomial" for a binary trait.

Value

A matrix containing the ranking score, the higher the more important.

Author(s)

Chong Wu

References

Chong, W., Pan, W. (2015) An Adaptive Association Test for Microbiome Data, submitted.

Examples

```
data(throat.otu.tab)
data(throat.tree)
data(throat.meta)

Y.tmp =throat.meta[,3]
Y = rep(0,dim(throat.meta)[1])
Y[Y.tmp=="Smoker"] = 1

cov.tmp = throat.meta[,c(10,12)]
cov = matrix(1,dim(throat.meta)[1],2)
cov[cov.tmp[,1]== "None",1] = 0
cov[cov.tmp[,2]== "Male",2] = 0

start.time = proc.time()
X = as.matrix(throat.otu.tab)

#out = MiSPU(Y,X, throat.tree,cov,model =  "binomial", pow = c(2:8, Inf), n.perm = 1000)
out = ranking(Y,X, throat.tree,cov,gamma = 2, g.taxon.index =1)
```

---

simulateData *OTU counts simulation*

Description

We used a phylogenetic tree of OTUs from a real throat microbiome data set, which consists of 856 OTUs after discarding singleton OTUs. Then based on a complicate statistical meodel, we generated the OTU counts for each individual to simulate the feature of real microbiome data.

Usage

```
simulateData(nSam=100, s=12, ncluster = 20, mu = 1000, size = 25)
```

Arguments

nSam	Sample size, the default value is 100.
s	The indicator of informative cluster.
ncluster	Positive integer specifying the number of clusters of they phylogenetic tree.
mu	The mean of the negative binomial distribution.
size	The size of the negative binomial distribution

Value

A list object, including the informative OTU counts table and whole OTU counts table.

Author(s)

Chong Wu

References

Chong, W., Pan, W. (2015) An Adaptive Association Test for Microbiome Data, submitted.

Examples

```
OTU = simulateData()
OTU
```

---

throat.meta *Meta data of the throat microbiome samples.*

Description

It is part of a microbiome data set for studying the effect of smoking on the upper respiratory tract microbiome. The original data set contains samples from both throat and nose microbiomes, and from both body sides. This data set comes from the throat microbiome of left body side. It contains 60 subjects consisting of 32 nonsmokers and 28 smokers.

Usage

```
data(throat.meta)
```

Format

The format is: chr "throat.meta"

Source

Charlson ES, Chen J, Custers-Allen R, Bittinger K, Li H, et al. (2010) Disordered Microbial Communities in the Upper Respiratory Tract of Cigarette Smokers. PLoS ONE 5(12): e15216.

Examples

```
data(throat.meta)
## maybe str(throat.meta) ; plot(throat.meta) ...
```

throat.otu.tab *OTU count table from 16S sequencing of the throat microbiome samples.*

Description

It is part of a microbiome data set for studying the effect of smoking on the upper respiratory tract microbiome. The original data set contains samples from both throat and nose microbiomes, and from both body sides. This data set comes from the throat microbiome of left body side. It contains 60 subjects consisting of 32 nonsmokers and 28 smokers.

Usage

```
data(throat.otu.tab)
```

Format

The format is: chr "throat.otu.tab"

Details

The OTU table is produced by the QIIME software. Singleton OTUs have been discarded.

Source

Charlson ES, Chen J, Custers-Allen R, Bittinger K, Li H, et al. (2010) Disordered Microbial Communities in the Upper Respiratory Tract of Cigarette Smokers. PLoS ONE 5(12): e15216.

Examples

```
data(throat.otu.tab)
## maybe str(throat.otu.tab) ; plot(throat.otu.tab) ...
```

throat.tree *UPGMA tree of the OTUs from 16S sequencing of the throat microbiome samples.*

Description

The OTU tree is constructed using UPGMA on the K80 distance matrice of the OTUs. It is a rooted tree of class "phylo".

Usage

```
data(throat.tree)
```

Format

The format is: chr "throat.tree"

Details

The OTUs are produced by the QIIME software. Singleton OTUs have been discarded.

Source

Charlson ES, Chen J, Custers-Allen R, Bittinger K, Li H, et al. (2010) Disordered Microbial Communities in the Upper Respiratory Tract of Cigarette Smokers. PLoS ONE 5(12): e15216.

Examples

```
data(throat.tree)
## maybe str(throat.tree) ; plot(throat.tree) ...
```

Index