

Package ‘SpotSampling’

October 26, 2020

Type Package

Title SPatial and Optimally Temporal (SPOT) Sampling

Version 0.1.0

Description In spatial data, information of two neighboring units are generally very similar. For spatial sampling, it is therefore more efficient to select samples that are well spread out in space. Often, the interest lies not only in estimating a measure at one point in time, but rather in estimating several points in time to also study the evolution. Three new methods called Orfs (Optimal Rotation with Fixed sample Size), Orsp (Optimal Rotation with Spread sample), and Spot (Spatial and Optimally Temporal Sampling) are implemented in this package. Orfs allows to select temporal samples with fixed size. Orsp selects spatio-temporal samples with random size that are well spread out in space at each point in time. And Spot generates spread sample with fixed sample size at each wave. These methods provide an optimal time rotation of the selected units using the systematic sampling.

License GPL (>= 2)

Encoding UTF-8

LazyData true

RoxygenNote 7.1.1

Depends R (>= 2.10)

Imports BalancedSampling, sampling, stats, utils, pracma,
WaveSampling, MASS

NeedsCompilation no

Author Esther Eustache [aut, cre],
Raphael Jauslin [aut] (<<https://orcid.org/0000-0003-1088-3356>>),
Yves Tille [aut] (<<https://orcid.org/0000-0003-0904-5523>>)

Maintainer Esther Eustache <esther.eustache@unine.ch>

Repository CRAN

Date/Publication 2020-10-26 10:40:06 UTC

R topics documented:

Orfs	2
Orsp	3
Preselection	4
ReducedMatrix	5
ReducedSamplecube	6
Spot	8
Spread	9
SystematicDesign	10
TemporalPivot	11

Index	13
--------------	-----------

Orfs	<i>ORFS method</i>
------	--------------------

Description

Select temporal samples with fixed size at each wave using the cube method. It provides optimal time rotation of the selected samples using the systematic sampling method.

Usage

```
Orfs(pik, EPS = 1e-08, comment = TRUE)
```

Arguments

pik	a matrix of temporal inclusion probabilities. Columns of pik correspond to the waves, and rows correspond to the units. Inclusion probabilities can be totally unequal.
EPS	a tolerance parameter. Default value is 1e-8.
comment	a comment is written during the execution if comment is TRUE (default value).

Value

a matrix that contains temporal samples. This is the update of pik and contains only 0s and 1s that indicates if a unit is selected or not at each wave.

Author(s)

Esther Eustache <esther.eustache@unine.ch>, Raphael Jauslin <raphael.jauslin@unine.ch>

References

Quenouille, M. H. (1949). Approximate Tests of Correlation in time-Series. Royal Statistical Society, Series B Vol. 11, No. 1 (1949), pp. 68-84.

See Also

[SystematicDesign](#), [ReducedSamplecube](#)

Examples

```
## Temporal inclusion probabilities with 3 waves and 4 units ##
pik <- matrix(c(0.6,0.3,0.3,
               0.2,0.4,0.9,
               0.3,0.2,0.5,
               0.9,0.1,0.3), ncol = 3, byrow = TRUE)
## ORFS method ##
Orfs(pik)
```

Orsp

ORSP method

Description

Select spatio-temporal samples with random size and well spread out in space at each wave. The pivotal method is used to obtain spread samples. It provides optimal time rotation of the selected sample using the systematic sampling method.

Usage

```
Orsp(pik, coord, EPS = 1e-06, comment = TRUE)
```

Arguments

pik	a matrix of temporal inclusion probabilities. Columns of pik correspond to the waves, and rows correspond to the units. Inclusion probabilities can be totally unequal.
coord	a matrix that contains spatial coordinates in columns. The number of columns can be more than two. Matrix rows correspond to the units.
EPS	a tolerance parameter. Default value is 1e-6.
comment	a comment is written during the execution if comment is TRUE (default value).

Value

a matrix that contains spatio-temporal samples. This is the update of pik and contains only 0s and 1s that indicates if a unit is selected or not at each wave.

Author(s)

Esther Eustache <esther.eustache@unine.ch>

References

- Quenouille, M. H. (1949). Approximate Tests of Correlation in time-Series. Royal Statistical Society, Series B Vol. 11, No. 1 (1949), pp. 68-84.
- Grafstrom, A., Lundstrom, N. L. P., and Schelin, L. (2012). Spatially balanced sampling through the pivotal method. *Biometrics*, 68(2):514-520.

See Also

[SystematicDesign](#), [lpm1](#)

Examples

```
## Coordinates in two dimensions of 4 units ##
coord <- matrix(c(0.5,0.6,0.2,0.3,0.8,0.9,0.4,0.7), ncol=2)
## Temporal inclusion probabilities with 3 waves and 4 units ##
pik <- matrix(c(0.6,0.3,0.3,
               0.2,0.4,0.9,
               0.3,0.2,0.5,
               0.9,0.1,0.3), ncol = 3, byrow = TRUE)
## ORSP method ##
Orsp(pik, coord, EPS = 1e-6)
```

Preselection

Selection of an initial spread set

Description

Select an initial spread set using the flight phase of the local cube method. Some inclusion probabilities are set to 0. The others probabilities are also modified so as not to change the sum of the inclusion probabilities.

Usage

```
Preselection(pik, coord, L = 1, EPS = 1e-09)
```

Arguments

pik	a matrix of temporal inclusion probabilities. Columns of pik correspond to the times, and rows correspond to the units. Inclusion probabilities can be totally unequal.
coord	a matrix that contains spatial coordinates in columns. The number of columns can be more than two. Matrix rows correspond to the units.
L	a parameter to achieve good spatial balanced (see details). Default value is 1.
EPS	a tolerance parameter. Default value is 1e-9.

Details

L is used to achieve good spatial balance. It must be equal to or larger than one.

Value

a matrix with the same size as pik that contains new temporal inclusion probabilities. Some inclusion probabilities are updated to 0.

Author(s)

Esther Eustache <esther.eustache@unine.ch>

References

Grafstrom, A., Lundstrom, N. L. P., and Schelin, L. (2012). Spatially balanced sampling through the pivotal method. *Biometrics*, 68(2):514-520.

See Also

[fastflightcube](#)

Examples

```
## Coordinates in two dimensions of 4 units ##
coord <- matrix(c(0.5,0.6,0.2,0.3,0.8,0.9,0.4,0.7), ncol=2)
## Temporal inclusion probabilities with 3 waves and 4 units ##
pik <- matrix(c(0.6,0.3,0.3,
               0.2,0.4,0.9,
               0.3,0.2,0.5,
               0.9,0.1,0.3), ncol = 3, byrow = TRUE)
## Selection of an initial spread set ##
Preselection(pik, coord)
```

ReducedMatrix

Matrix reduction

Description

Reduce a matrix by removing alternatively columns and rows that sum to 0. If the matrix is dense or if every columns sum to more than 0, then nothing is changed.

Usage

```
ReducedMatrix(B)
```

Arguments

B a matrix that contains lot of 0s.

Value

Returns a list including:

BB the reduced matrix of B.

ind_col a vector that contains the index of the remaining columns of B in BB.

ind_row a vector that contains the index of the remaining rows of B in BB.

Author(s)

Raphael Jauslin <raphael.jauslin@unine.ch>

Examples

```
set.seed(1)
B <- matrix(sample(c(0,0,0,1),80,replace=TRUE), nrow = 8, ncol = 10)
ReducedMatrix(B)
```

ReducedSamplecube

Cube method with reduction of the auxiliary variables matrix

Description

Modified cube method. This function reduces considerably the execution time when the matrix of auxiliary variables X contains lot of 0s. It is based on the function [samplecube](#) from the package `sampling`.

Usage

```
ReducedSamplecube(X, pik, redux = TRUE, t)
```

Arguments

<code>X</code>	a matrix of size $(N \times p)$ of auxiliary variables on which the sample must be balanced.
<code>pik</code>	a vector of size N of inclusion probabilities.
<code>redux</code>	a boolean value that specify if matrix X is reduced during the cube method. Default value is TRUE.
<code>t</code>	the maximum number of constraints that can potentially be removed during the landing phase.

Details

In case where the number of auxiliary variables is great (i.e. p very large), even if we use the fast implementation proposed by (Chauvet and Tille 2005), the problem is time consuming. This function reduces considerably the execution time when the matrix of auxiliary variables X contains lot of 0s. It considers a reduced matrix X by removing columns and rows that sum to 0 (see [ReducedMatrix](#)).

Moreover, the landing by variable suppression is used. t specifies the maximum number of constraints that can potentially be removed during the landing phase. This means that the first $(N-T)$ constraints in X can be exactly satisfied.

Value

the updated vector of pik that contains only 0s and 1s that indicates if a unit is selected or not at each wave.

Author(s)

Esther Eustache <esther.eustache@unine.ch>, Raphael Jauslin <raphael.jauslin@unine.ch>

References

Chauvet, G. and Tille, Y. (2006). A fast algorithm of balanced sampling. *Computational Statistics*, 21/1:53-62

See Also

[samplecube](#), [landingcube](#), [ReducedMatrix](#).

Examples

```
set.seed(1)
## Matrix of 8 auxiliary variables and 10 units with lot of 0s ##
X <- matrix(c(0.6,0.0,0.0,0.0,
             0.1,0.0,0.1,0.0,
             0.3,0.0,0.0,0.3,
             0.0,0.3,0.0,0.3,
             0.0,0.6,0.0,0.0,
             0.0,0.1,0.1,0.0), ncol = 4, byrow = TRUE)

## Inclusion probabilities with 10 units ##
pik <- c(0.60,0.10,0.30,0.30,0.60,0.10)

## parameter t ##
t <- 2

## Cube method ##
s <- ReducedSamplecube(X, pik, redux = TRUE, t)
s
```

Spot

SPOT method

Description

Select spatio-temporal samples of fixed size and well spread out in space at each wave. The pivotal method is used to obtain spread samples. It provides optimal time rotation of the selected sample using the systematic sampling method.

Usage

```
Spot(pik, coord, EPS = 1e-08, comment = TRUE)
```

Arguments

pik	a matrix of temporal inclusion probabilities. Columns of pik correspond to the waves, and rows correspond to the units. Inclusion probabilities can be totally unequal.
coord	a matrix that contains spatial coordinates in columns. The number of columns can be more than two. Matrix rows correspond to the units.
EPS	a tolerance parameter. Default value is 1e-8.
comment	a comment is written during the execution if comment is TRUE (default value).

Value

a matrix that contains temporal samples. This is the update of pik and contains only 0s and 1s that indicates if a unit is selected or not at each wave.

Author(s)

Esther Eustache <esther.eustache@unine.ch>

References

Quenouille, M. H. (1949). Approximate Tests of Correlation in time-Series. Royal Statistical Society, Series B Vol. 11, No. 1 (1949), pp. 68-84.

Grafstrom, A., Lundstrom, N. L. P., and Schelin, L. (2012). Spatially balanced sampling through the pivotal method. Biometrics, 68(2):514-520.

See Also

[SystematicDesign](#), [ReducedSamplecube](#).

Examples

```
## Coordinates in two dimensions of 4 units ##
coord <- matrix(c(0.5,0.6,0.2,0.3,0.8,0.9,0.4,0.7), ncol=2)
## Temporal inclusion probabilities with 3 waves and 4 units ##
pik <- matrix(c(0.6,0.3,0.3,
               0.2,0.4,0.9,
               0.3,0.2,0.5,
               0.9,0.1,0.3), ncol = 3, byrow = TRUE)
## SPOT method ##
Spot(pik, coord, EPS = 1e-6)
```

Spread

Spreading measure of several spatial samples

Description

Measure the spread of several spatial samples depending on inclusion probabilities and spatial coordinates. Two spreading criteria are available: one based on the space partition of Voronoi polygons and one based on Moran's I index (see references).

Usage

```
Spread(S, pik, coord, criteria)
```

Arguments

S	a matrix that contains samples in columns. Matrix rows correspond to the units. It could be matrix of temporal samples returned by function Spot .
pik	a matrix of temporal inclusion probabilities. Columns of pik correspond to samples, and rows correspond to the units.
coord	a matrix that contains spatial coordinates in columns. Matrix rows correspond to the units.
criteria	it specifies the criteria used to measure samples spreading. <code>criteria = "IB"</code> : the criteria based on Moran's I index is used (see IB), <code>criteria = "sb"</code> : the criteria based on Moran's I index is used (see sb).

Value

a vector that contains values of the spreading measure of the samples in columns of S.

Author(s)

Esther Eustache <esther.eustache@unine.ch>

References

Grafstrom, A., Lundstrom, N. L. P., and Schelin, L. (2012). Spatially balanced sampling through the pivotal method. *Biometrics*, 68(2):514-520.

Jauslin, R. and Tille, Y. (2019). Spatial spread sampling using weakly associated vectors. Statistical Office, University of Neuchatel.

See Also

[IB](#), [sb](#).

Examples

```
set.seed(1)
## Coordinates in two dimensions of 10 units ##
coord <- matrix(stats::runif(10*2), ncol=2)
## Temporal inclusion probabilities with 3 waves and 4 units ##
pik <- matrix(rep(0.2,10*3), ncol = 3, byrow = TRUE)
## Spot method to obtain temporal samples ##
S <- Spot(pik, coord)
## Compute IB criteria ##
Spread(S, pik, coord, criteria = 'IB')
```

SystematicDesign

Systematic sampling design

Description

Find the systematic sampling design of an inclusion probabilities vector. It gives all possible samples and their selection probabilities using the systematic sampling.

Usage

```
SystematicDesign(pik, EPS = 1e-06)
```

Arguments

<code>pik</code>	a vector of inclusion probabilities.
<code>EPS</code>	a tolerance parameter. Default value is 1e-6.

Value

Returns a list including:

`samples` a matrix that contains the systematic samples in rows. The samples have the same length as vector `pik` and contains 0s and 1s that specify whether a unit is selected or not in the sample.

`probas` a vector that contains the selection probabilities of samples in `samples`.

Author(s)

Esther Eustache <esther.eustache@unine.ch>

References

Quenouille, M. H. (1949). Approximate Tests of Correlation in time-Series. Royal Statistical Society, Series B Vol. 11, No. 1 (1949), pp. 68-84.

Tille, Y. (2020). Sampling and Estimation from Finite Populations. John Wiley & Sons, 91(4), page 89.

Examples

```
## Vector of inclusion probabilities ##
pik  <- c(0.2,0.3,0.6)
## Find the systematic sampling design of pik ##
SystematicDesign(pik)
```

TemporalPivot

Adaptation of the local pivotal method on temporal samples

Description

This function considers longitudinal systematic sampling designs of two different units that result from function [SystematicDesign](#). It allows to decide if one of these 2 units is selected at a specific time by putting at least one of the samples selection probabilities to 0. It is based on the local pivotal method.

Usage

```
TemporalPivot(design1, design2, d, EPS = 1e-06)
```

Arguments

design1	a longitudinal systematic sampling design of a first unit. The length of the longitudinal samples is T . It results from function SystematicDesign .
design2	a longitudinal systematic sampling design of a second unit. The length of the longitudinal samples is T . It results from function SystematicDesign .
d	a vector of size T that specify for which time t a decision must be taken, with $1 \leq t \leq T$. d is such that the t -th element is equal to 1, and the others to 0.
EPS	a tolerance parameter. Default value is 1e-6.

Value

Returns a list including:

p1_new the updated probabilities of the longitudinal systematic sampling design of the first unit.

p2_new the updated probabilities of the longitudinal systematic sampling design of the second unit.

Author(s)

Esther Eustache <esther.eustache@unine.ch>

References

Quenouille, M. H. (1949). Approximate Tests of Correlation in time-Series. Royal Statistical Society, Series B Vol. 11, No. 1 (1949), pp. 68-84.

Tille, Y. (2020). Sampling and Estimation from Finite Populations. John Wiley & Sons, 91(4), page 89.

Grafstrom, A., Lundstrom, N. L. P., and Schelin, L. (2012). Spatially balanced sampling through the pivotal method. Biometrics, 68(2):514-520.

See Also

[SystematicDesign](#)

Examples

```
## Vectors of temporal inclusion probabilities with 3 waves ##
pik1  <- c(0.2,0.3,0.5) # of a first unit
pik2  <- c(0.1,0.4,0.5) # of a second unit
## Find the systematic sampling designs of pik1 and pik2 ##
design1 <- SystematicDesign(pik1, EPS = 1e-6)
design2 <- SystematicDesign(pik2, EPS = 1e-6)
## The time we want to take a decision ##
t     <- 2
d     <- rep(0,3)
d[t] <- 1
## Update probabilities to take a decision at wave t=2 ##
TemporalPivot(design1, design2, d)
```

Index

fastflightcube, 5

IB, 9, 10

landingcube, 7

lpm1, 4

Orfs, 2

Orsp, 3

Preselection, 4

ReducedMatrix, 5, 7

ReducedSamplecube, 3, 6, 8

samplecube, 6, 7

sb, 9, 10

Spot, 8, 9

Spread, 9

SystematicDesign, 3, 4, 8, 10, 11, 12

TemporalPivot, 11