Package ‘infotheo’

October 13, 2022

**Title**  Information-Theoretic Measures

**Version**  1.2.0.1

**Date**  2014-07-16

**Author**  Patrick E. Meyer

**Description**  Implements various measures of information theory based on several entropy estimators.

**Maintainer**  Patrick E. Meyer <software@meyerp.com>

**License**  GPL (>= 3)

**URL**  [http://homepage.meyerp.com/software](http://homepage.meyerp.com/software)

**Repository**  CRAN

**NeedsCompilation**  yes

**Date/Publication**  2022-04-08 11:00:24 UTC

---

**R topics documented:**

condentropy ................................................................. 2  
condinformation .......................................................... 3  
discretize ................................................................. 4  
entropy ........................................................................... 5  
infotheo ................................................................. 6  
interinformation .......................................................... 7  
multiinformation ......................................................... 8  
mutinformation ............................................................ 9  
natstobits ............................................................... 10

**Index**  12

1
condentropy

**condentropy**

**condentropy** takes two random vectors, X and Y, as input and returns the conditional entropy, H(X|Y), in nats (base e), according to the entropy estimator method. If Y is not supplied the function returns the entropy of X - see **entropy**.

**Usage**

```r
condentropy(X, Y=NULL, method="emp")
```

**Arguments**

- **X**: data.frame denoting a random variable or random vector where columns contain variables/features and rows contain outcomes/samples.
- **Y**: data.frame denoting a conditioning random variable or random vector where columns contain variables/features and rows contain outcomes/samples.
- **method**: The name of the entropy estimator. The package implements four estimators: "emp", "mm", "shrink", "sg" (default:"emp") - see details. These estimators require discrete data values - see **discretize**.

**Details**

- "emp" : This estimator computes the entropy of the empirical probability distribution.
- "mm" : This is the Miller-Madow asymptotic bias corrected empirical estimator.
- "shrink" : This is a shrinkage estimate of the entropy of a Dirichlet probability distribution.
- "sg" : This is the Schurmann-Grassberger estimate of the entropy of a Dirichlet probability distribution.

**Value**

condentropy returns the conditional entropy, H(X|Y), of X given Y in nats.

**Author(s)**

Patrick E. Meyer

**References**


condinformation

See Also

entropy, mutinformation, natstobits

Examples

data(USArrests)
dat<-discretize(USArrests)
H <- condentropy(dat[,1], dat[,2], method = "mm")

condinformation conditional mutual information computation

Description

condinformation takes three random variables as input and computes the conditional mutual information in nats according to the entropy estimator method. If S is not supplied the function returns the mutual information between X and Y - see mutinformation

Usage

condinformation(X, Y, S=NULL, method="emp")

Arguments

X vector/factor denoting a random variable or a data.frame denoting a random vector where columns contain variables/features and rows contain outcomes/samples.
Y another random variable or random vector (vector/factor or data.frame).
S the conditioning random variable or random vector (vector/factor or data.frame).
method The name of the entropy estimator. The package implements four estimators: "emp", "mm", "shrink", "sg" (default:"emp") - see details. These estimators require discrete data values - see discretize.

Details

• "emp" : This estimator computes the entropy of the empirical probability distribution.
• "mm" : This is the Miller-Madow asymptotic bias corrected empirical estimator.
• "shrink" : This is a shrinkage estimate of the entropy of a Dirichlet probability distribution.
• "sg" : This is the Schurmann-Grassberger estimate of the entropy of a Dirichlet probability distribution.

Value

condinformation returns the conditional mutual information, I(X;Y|S), in nats.

Author(s)

Patrick E. Meyer
References


See Also

mutinformation, multiinformation, interinformation, natstobits

Examples

data(USArrests)
dat<-discretize(USArrests)
I <- condinformation(dat[,1],dat[,2],dat[,3],method="emp")

---

discretize

Unsupervised Data Discretization

Description

discretize discretizes data using the equal frequencies or equal width binning algorithm. "equalwidth" and "equalfreq" discretizes each random variable (each column) of the data into nbins. "globalequalwidth" discretizes the range of the random vector data into nbins.

Usage

discretize( X, disc="equalfreq", nbins=NROW(X)^(1/3) )

Arguments

X A data.frame containing data to be discretized. The columns contains variables and the rows samples.

disc The name of the discretization method to be used :"equalfreq", "equalwidth" or "globalequalwidth" (default : "equalfreq") - see references.

nbins Integer specifying the number of bins to be used for the discretization. By default the number of bins is set to N^{1/3} where N is the number of samples.

Value

discretize returns the discretized dataset.

Author(s)

Patrick E. Meyer, Frederic Lafitte, Gianluca Bontempi, Korbinian Strimmer
entropy

References


Examples
```r
data(USArrests)
nbins <- sqrt(NROW(USArrests))
ew.data <- discretize(USArrests,"equalwidth", nbins)
eref.data <- discretize(USArrests,"equalfreq", nbins)
gew.data <- discretize(USArrests,"globalequalwidth", nbins)
```

Description
entropy takes the dataset as input and computes the entropy according to the entropy estimator method.

Usage
```r
entropy(X, method="emp")
```

Arguments
- **X**
  - data.frame denoting a random vector where columns contain variables/features and rows contain outcomes/samples.
- **method**
  - The name of the entropy estimator. The package implements four estimators: "emp", "mm", "shrink", "sg" (default: "emp") - see details. These estimators require discrete data values - see discretize.

Details
- "emp" : This estimator computes the entropy of the empirical probability distribution.
- "mm" : This is the Miller-Madow asymptotic bias corrected empirical estimator.
- "shrink" : This is a shrinkage estimate of the entropy of a Dirichlet probability distribution.
- "sg" : This is the Schurmann-Grassberger estimate of the entropy of a Dirichlet probability distribution.
Value

entropy returns the entropy of the data in nats.

Author(s)

Patrick E. Meyer

References


See Also

condentropy, mutinformation, natstobits

Examples

data(USArrests)
H <- entropy(discretize(USArrests), method="shrink")
interinformation takes a dataset as input and computes the interaction information among the random variables in the dataset using the entropy estimator method. This measure is also called synergy or complementarity.

Usage

interinformation(X, method="emp")

Arguments

- **X**: data.frame denoting a random vector where columns contain variables/features and rows contain outcomes/samples.
- **method**: The name of the entropy estimator. The package implements four estimators: "emp", "mm", "shrink", "sg" (default:"emp") - see details. These estimators require discrete data values - see discretize.

Details

- "emp": This estimator computes the entropy of the empirical probability distribution.
- "mm": This is the Miller-Madow asymptotic bias corrected empirical estimator.
- "shrink": This is a shrinkage estimate of the entropy of a Dirichlet probability distribution.
- "sg": This is the Schurmann-Grassberger estimate of the entropy of a Dirichlet probability distribution.

Value

interinformation returns the interaction information (also called synergy or complementarity), in nats, among the random variables (columns of the data.frame).

Author(s)

Patrick E. Meyer

References


multiinformation

See Also
condinformation, multiinformation, mutinformation, natstobits

Examples

```r
data(USArrests)
dat<-discretize(USArrests)
ii <- interinformation(dat, method = "sg")
```

Description

multiinformation takes a dataset as input and computes the multiinformation (also called total correlation) among the random variables in the dataset. The value is returned in nats using the entropy estimator estimator.

Usage

```r
multiinformation(X, method = "emp")
```

Arguments

- **X**: data.frame containing a set of random variables where columns contain variables/features and rows contain outcomes/samples.
- **method**: The name of the entropy estimator. The package implements four estimators: "emp", "mm", "shrink", "sg" (default:"emp") - see details. These estimators require discrete data values - see discretize.

Details

- "emp": This estimator computes the entropy of the empirical probability distribution.
- "mm": This is the Miller-Madow asymptotic bias corrected empirical estimator.
- "shrink": This is a shrinkage estimate of the entropy of a Dirichlet probability distribution.
- "sg": This is the Schurmann-Grassberger estimate of the entropy of a Dirichlet probability distribution.

Value

multiinformation returns the multiinformation (also called total correlation) among the variables in the dataset (in nats).

Author(s)

Patrick E. Meyer
mutinformation

References


See Also
condinformation, mutinformation, interinformation, natstobits

Examples

data(USArrests)
dat<-discretize(USArrests)
M <- multiinformation(dat)

Description

mutinformation takes two random variables as input and computes the mutual information in nats according to the entropy estimator method. If Y is not supplied and X is a matrix-like argument, the function returns a matrix of mutual information between all pairs of variables in the dataset X.

Usage

mutinformation(X, Y, method="emp")

Arguments

X  vector/factor denoting a random variable or a data.frame denoting a random vector where columns contain variables/features and rows contain outcomes/samples.

Y  another random variable or random vector (vector/factor or data.frame).

method  The name of the entropy estimator. The package implements four estimators: "emp", "mm", "shrink", "sg" (default:"emp") - see details. These estimators require discrete data values - see discretize.

Details

- "emp": This estimator computes the entropy of the empirical probability distribution.
- "mm": This is the Miller-Madow asymptotic bias corrected empirical estimator.
- "shrink": This is a shrinkage estimate of the entropy of a Dirichlet probability distribution.
- "sg": This is the Schurmann-Grassberger estimate of the entropy of a Dirichlet probability distribution.
Value

mutinformation returns the mutual information $I(X;Y)$ in nats.

Author(s)

Patrick E. Meyer

References


See Also

condinformation, multiinformation, interinformation, natstobits

Examples

data(USArrests)
dat<-discretize(USArrests)
#computes the MIM (mutual information matrix)
I <- mutinformation(dat, method = "emp")
I2 <- mutinformation(dat[,1], dat[,2])

natstobits

convert nats into bits

Description

natstobits takes a value in nats (a double) as input and returns the value in bits (a double).

Usage

natstobits(H)

Arguments

H double denoting a value (in nats), as returned by one of the function of the infotheo package

Details

Information-theoretic quantities can have different units depending on the base of the logarithm used in their computation. All the function of the package use a base $e$, hence the unit is the nat. The value in bit is given by using the base 2, hence the conversion is done by multiplying by $\log_2(e) = 1.442695$. 
Value

natstobits returns a double that is the conversion of the nats value into bits.

Author(s)

Patrick E. Meyer

Examples

```r
data(USArrests)
H <- entropy(discretize(USArrests))
natstobits(H)
```
Index

* misc
  condentropy, 2
  condinformation, 3
  discretize, 4
  entropy, 5
  infotheo, 6
  interinformation, 7
  multiinformation, 8
  mutinformation, 9
  natstobits, 10

condentropy, 2, 6
condinformation, 3, 6, 8–10

discretize, 2, 3, 4, 5, 7–9

entropy, 2, 3, 5, 6

infotheo, 6
interinformation, 4, 6, 7, 9, 10

multiinformation, 4, 6, 8, 8, 10
mutinformation, 3, 4, 6, 8, 9, 9

natstobits, 3, 4, 6, 8–10, 10