Package 'EntropyEstimation'

September 15, 2024

Contents

EntropyEstimation-package

Estimation of Entropy and Related Quantities

Description

Contains methods for the estimation of Shannon's entropy, variants of Renyi's entropy, mutual Information, Kullback-Leibler divergence, and generalized Simpson's indices. These estimators have a bias that decays exponentially fast. For more information see Z. Zhang and J. Zhou (2010), Zhang (2012), Zhang (2013), Zhang and Grabchak (2013), Zhang and Grabchak (2014a), Zhang and Grabchak (2014b), and Zhang and Zheng (2014).

Details

Package:	EntropyEstimation
Type:	Package
Version:	1.2.1
Date:	2024-09-14
License:	GPL3

Author(s)

Lijuan Cao <lcao2@charlotte.edu> and Michael Grabchak <mgrabcha@charlotte.edu>

References

Z. Zhang (2012). Entropy estimation in Turing's' perspective. Neural Computation 24(5), 1368–1389.

Z. Zhang (2013). Asymptotic normality of an entropy estimator with asymptotically decaying bias. IEEE Transactions on Information Theory 59(1), 504–508.

Z. Zhang and M. Grabchak (2013). Bias Adjustment for a Nonparametric Entropy Estimator. Entropy, 15(6), 1999-2011.

Z. Zhang and M. Grabchak (2014a). Entropic representation and estimation of diversity indices. http://arxiv.org/abs/1403.3031.

Z. Zhang and M. Grabchak (2014b). Nonparametric Estimation of Kullback-Leibler Divergence. Neural Computation, 26(11): 2570-2593.

Z. Zhang and L. Zheng (2014). A Mutual Information Estimator with Exponentially Decaying Bias.

Z. Zhang and J. Zhou (2010). Re-parameterization of multinomial distributions and diversity indices. Journal of Statistical Planning and Inference 140(7), 1731-1738.

Entropy.sd   *Entropy.sd*

Description

Returns the estimated asymptotic standard deviation for the Z estimator of Shannon's Entropy. Note that this is also the asymptotic standard deviation of the plug-in estimator. See Zhang and Grabchak (2014a) for details.

Usage

```
Entropy.sd(x)
```

Arguments

x   Vector of counts. Must be integer valued. Each entry represents the number of observations of a distinct letter.

Author(s)

Lijuan Cao and Michael Grabchak

References

Z. Zhang and M. Grabchak (2014a). Entropic representation and estimation of diversity indices. http://arxiv.org/abs/1403.3031.

Examples

```
x = c(1,3,7,4,8) # vector of counts
Entropy.sd(x)  # Estimated standard deviation

data = rbinom(10,20,.5)
counts = tabulate(as.factor(data))
Entropy.sd(counts)
```

Entropy.z	*Entropy.z*

Description

Returns the Z estimator of Shannon's Entropy. This estimator has exponentially decaying bias. See Zhang (2012), Zhang (2013), and Zhang and Grabchak (2014a) for details.

Usage

```
Entropy.z(x)
```

Arguments

x   Vector of counts. Must be integer valued. Each entry represents the number of observations of a distinct letter.

Author(s)

Lijuan Cao and Michael Grabchak

References

Z. Zhang (2012). Entropy estimation in Turing's' perspective. Neural Computation 24(5), 1368–1389.

Z. Zhang (2013). Asymptotic normality of an entropy estimator with asymptotically decaying bias. IEEE Transactions on Information Theory 59(1), 504–508.

Z. Zhang and M. Grabchak (2014a). Entropic representation and estimation of diversity indices. http://arxiv.org/abs/1403.3031.

Examples

```
x = c(1,3,7,4,8)
Entropy.z(x)

data = rbinom(10,20,.5)
counts = tabulate(as.factor(data))
Entropy.z(counts)
```

GenSimp.sd	*GenSimp.sd*

Description

Returns the estimated asymptotic standard deviation of the Z estimator of the generalized Simpson's index of order r, i.e. of the index sum_k p_k(1-p_k)^r. This estimate of the standard deviation is based on the formula in Zhang and Grabchak (2014a) and not the one in Zhang and Zhou (2010).

Usage

```
GenSimp.sd(x, r)
```

Arguments

x   Vector of counts. Must be integer valued. Each entry represents the number of observations of a distinct letter.

r   Positive integer representing the order of the generalized Simpson's index. If a noninteger value is given then the integer part is taken. Must be strictly less than sum(x).

Author(s)

Lijuan Cao and Michael Grabchak

References

Z. Zhang and M. Grabchak (2014a). Entropic representation and estimation of diversity indices. http://arxiv.org/abs/1403.3031.

Z. Zhang and J. Zhou (2010). Re-parameterization of multinomial distributions and diversity indices. Journal of Statistical Planning and Inference 140(7), 1731-1738.

Examples

```
x = c(1,3,7,4,8)
GenSimp.sd(x,2)

data = rbinom(10,20,.5)
counts = tabulate(as.factor(data))
GenSimp.sd(counts,2)
```

GenSimp.z	*GenSimp.z*

Description

Returns the Z estimator of the generalized Simpson's index of order r, i.e. of the index sum_k p_k(1-p_k)^r. See Zhang and Zhou (2010) and Zhang and Grabchak (2014a) for details.

Usage

```
GenSimp.z(x,r)
```

Arguments

x	Vector of counts. Must be integer valued. Each entry represents the number of observations of a distinct letter.
r	Positive integer representing the order of the generalized Simpson's index. If a noninteger value is given then the integer part is taken. Must be strictly less than sum(x).

Author(s)

Lijuan Cao and Michael Grabchak

References

Z. Zhang and M. Grabchak (2014a). Entropic representation and estimation of diversity indices. http://arxiv.org/abs/1403.3031.

Z. Zhang and J. Zhou (2010). Re-parameterization of multinomial distributions and diversity indices. Journal of Statistical Planning and Inference 140(7), 1731-1738.

Examples

```
x = c(1,3,7,4,8)
GenSimp.z(x,2)

data = rbinom(10,20,.5)
counts = tabulate(as.factor(data))
GenSimp.z(counts,2)
```

`Hill.sd` *Hill.sd*

Description

Returns the estimated asymptotic standard deviation for the Z estimator of Hill's diversity numbe. Note that this is also the asymptotic standard deviation of the plug-in estimator. See Zhang and Grabchak (2014a) for details.

Usage

```
Hill.sd(x, r)
```

Arguments

x	Vector of counts. Must be integer valued. Each entry represents the number of observations of a distinct letter.
r	Order of Hill's deversity numbe. Must be a strictly positive real number.

Author(s)

Lijuan Cao and Michael Grabchak

References

Z. Zhang and M. Grabchak (2014a). Entropic representation and estimation of diversity indices. http://arxiv.org/abs/1403.3031.

Examples

```
x = c(1,3,7,4,8)
Hill.sd(x,2)

data = rbinom(10,20,.5)
counts = tabulate(as.factor(data))
Hill.sd(counts,2)
```

`Hill.z` *Hill.z*

Description

Returns the Z estimator of Hill's diversity number. This is based on raising the Z estimator of Renyi's equivalent entropy to the 1/(r-1) power. When r=1 returns exp(H), where H is the Z estimator of Shannon's entropy. See Zhang and Grabchak (2014a) for details.

Usage

```
Hill.z(x, r)
```

Arguments

x	Vector of counts. Must be integer valued. Each entry represents the number of observations of a distinct letter.
r	Order of Renyi's equivalent entropy this index is based on. Must be a strictly positive real number.

Author(s)

Lijuan Cao and Michael Grabchak

References

Z. Zhang and M. Grabchak (2014a). Entropic representation and estimation of diversity indices. http://arxiv.org/abs/1403.3031.

Examples

```
x = c(1,3,7,4,8)
Hill.z(x,2)

data = rbinom(10,20,.5)
counts = tabulate(as.factor(data))
Hill.z(counts,2)
```

KL.Plugin   *KL.Plugin*

Description

Returns the augmented plugin estimator of Kullback-Leibler Divergence. See Zhang and Grabchak (2014b) for details.

Usage

```
KL.Plugin(x, y)
```

Arguments

x	Vector of counts from first distribution. Must be integer valued. Each entry represents the number of observations of a distinct letter.
y	Vector of counts from second distribution. Must be integer valued. Each entry represents the number of observations of a distinct letter.

Author(s)

Lijuan Cao and Michael Grabchak

References

Z. Zhang and M. Grabchak (2014b). Nonparametric Estimation of Kullback-Leibler Divergence. Neural Computation, 26(11): 2570-2593.

Examples

```
x = c(1,3,7,4,8)
y = c(2,5,1,3,6)
KL.Plugin(x,y)
KL.Plugin(y,x)
```

KL.sd	*KL.sd*

Description

Returns the estimated asymptotic standard deviation for the Z estimator of Kullback-Leibler's divergence. Note that this is also the asymptotic standard deviation of the plug-in estimator. See Zhang and Grabchak (2014b) for details.

Usage

```
KL.sd(x, y)
```

Arguments

x	Vector of counts from the first distribution. Must be integer valued. Each entry represents the number of observations of a distinct letter.
y	Vector of counts from the second distribution. Must be integer valued. Each entry represents the number of observations of a distinct letter.

Author(s)

Lijuan Cao and Michael Grabchak

References

Z. Zhang and M. Grabchak (2014b). Nonparametric Estimation of Kullback-Leibler Divergence. Neural Computation, 26(11): 2570-2593.

Examples

```
x = c(1,3,7,4,8) # first vector of counts
y = c(2,5,1,3,6) # second vector of counts
KL.sd(x,y)  # Estimated standard deviation
KL.sd(y,x)  # Estimated standard deviation
```

KL.z   *KL.z*

Description

Returns the Z estimator of Kullback-Leibler Divergence, which has exponentially decaying bias. See Zhang and Grabchak (2014b) for details.

Usage

```
KL.z(x, y)
```

Arguments

x	Vector of counts from the first distribution. Must be integer valued. Each entry represents the number of observations of a distinct letter.
y	Vector of counts from the second distribution. Must be integer valued. Each entry represents the number of observations of a distinct letter.

Author(s)

Lijuan Cao and Michael Grabchak

References

Z. Zhang and M. Grabchak (2014b). Nonparametric Estimation of Kullback-Leibler Divergence. Neural Computation, 26(11): 2570-2593.

Examples

```
x = c(1,3,7,4,8)
y = c(2,5,1,3,6)
KL.z(x,y)
KL.z(y,x)
```

MI.sd   *MI.sd*

Description

Returns the estimated asymptotic standard deviation for the Z estimator of mutual information. Note that this is also the asymptotic standard deviation of the plug-in estimator. See Zhang and Zheng (2014) for details.

Usage

```
MI.sd(y)
```

Arguments

y   Matrix of counts. Must be integer valued. Each entry represents the number of observations of a distinct combination of letters from the two alphabets.

Author(s)

Lijuan Cao and Michael Grabchak

References

Z. Zhang and L. Zheng (2014). A Mutual Information Estimator with Exponentially Decaying Bias.

Examples

```
 x = matrix(c(0, 0, 0, 1, 1, 0, 0, 0, 0, 0,
       0, 0, 0, 1, 0, 0, 1, 1, 0, 1,
       0, 0, 0, 2, 1, 0, 1, 0, 0, 1,
       0, 0, 0, 1, 1, 2, 0, 0, 0, 0,
       0, 0, 0, 3, 6, 2, 2, 0, 0, 0,
       2, 0, 2, 5, 6, 5, 1, 0, 0, 0,
       0, 0, 4, 6, 11, 5, 1, 1, 0, 1,
       0, 0, 5, 10, 21, 7, 5, 1, 0, 1,
       0, 0, 7, 11, 9, 6, 3, 0, 0, 1,
       0, 0, 4, 10, 6, 5, 1, 0, 0, 0),10,10,byrow=TRUE)
MI.sd(x)

x = rbinom(100,20,.5)
y = rbinom(100,20,.5)
MI.sd(table(x,y))
```

MI.z   *MI.z*

Description

Returns the Z estimator of Mutual Information. This estimator has exponentially decaying bias. See Zhang and Zheng (2014) for details.

Usage

```
MI.z(x)
```

Arguments

x   Matrix of counts. Must be integer valued. Each entry represents the number of observations of a distinct combination of letters from the two alphabets.

Author(s)

Lijuan Cao and Michael Grabchak

References

Z. Zhang and L. Zheng (2014). A Mutual Information Estimator with Exponentially Decaying Bias.

Examples

```
x = matrix(c(0, 0, 0, 1, 1, 0, 0, 0, 0, 0,
       0, 0, 0, 1, 0, 0, 1, 1, 0, 1,
       0, 0, 0, 2, 1, 0, 1, 0, 0, 1,
       0, 0, 0, 1, 1, 2, 0, 0, 0, 0,
       0, 0, 0, 3, 6, 2, 2, 0, 0, 0,
       2, 0, 2, 5, 6, 5, 1, 0, 0, 0,
       0, 0, 4, 6, 11, 5, 1, 1, 0, 1,
       0, 0, 5, 10, 21, 7, 5, 1, 0, 1,
       0, 0, 7, 11, 9, 6, 3, 0, 0, 1,
       0, 0, 4, 10, 6, 5, 1, 0, 0, 0),10,10,byrow=TRUE)
MI.z(x)


x = rbinom(100,20,.5)
y = rbinom(100,20,.5)
MI.z(table(x,y))
```

Renyi.sd *Renyi.sd*

Description

Returns the estimated asymptotic standard deviation for the Z estimator of Renyi's Entropy. Note that this is also the asymptotic standard deviation of the plug-in estimator. See Zhang and Grabchak (2014a) for details.

Usage

```
Renyi.sd(x, r)
```

Arguments

x   Vector of counts. Must be integer valued. Each entry represents the number of observations of a distinct letter.

r   Order of Renyi's entropy. Must be a strictly positive real number. Not allowed to be 1, in that case use Entropy.sd instead.

Author(s)

Lijuan Cao and Michael Grabchak

References

Z. Zhang and M. Grabchak (2014a). Entropic representation and estimation of diversity indices. http://arxiv.org/abs/1403.3031.

Examples

```
x = c(1,3,7,4,8)
Renyi.sd(x,2)

data = rbinom(10,20,.5)
counts = tabulate(as.factor(data))
Renyi.sd(counts,2)
```

Renyi.z   *Renyi.z*

Description

Returns the Z estimator of Renyi's entropy. This is based on taking the log of the Z estimator of Renyi's equivalent entropy and dividing by (1-r). When r=1 returns the Z estimator of Shannon's entropy. See Zhang and Grabchak (2014a) for details.

Usage

```
Renyi.z(x, r)
```

Arguments

x	Vector of counts. Must be integer valued. Each entry represents the number of observations of a distinct letter.
r	Order of Renyi's equivalent entropy this index is based on. Must be a strictly positive real number.

Author(s)

Lijuan Cao and Michael Grabchak

References

Z. Zhang and M. Grabchak (2014a). Entropic representation and estimation of diversity indices. http://arxiv.org/abs/1403.3031.

Examples

```
x = c(1,3,7,4,8)
Renyi.z(x,2)

data = rbinom(10,20,.5)
counts = tabulate(as.factor(data))
Renyi.z(counts,2)
```

RenyiEq.sd   *RenyiEq.sd*

Description

Returns the estimated asymptotic standard deviation for the Z estimator of Renyi Equivalent Entropy. Note that this is also the asymptotic standard deviation of the plug-in estimator. When r=1, returns 0. See Zhang and Grabchak (2014a) for details.

Usage

```
RenyiEq.sd(x, r)
```

Arguments

x   Vector of counts. Must be integer valued. Each entry represents the number of observations of a distinct letter.

r   Order of Renyi's equivalent entropy. Must be a strictly positive real number.

Author(s)

Lijuan Cao and Michael Grabchak

References

Z. Zhang and M. Grabchak (2014a). Entropic representation and estimation of diversity indices. http://arxiv.org/abs/1403.3031.

Examples

```
x = c(1,3,7,4,8)
RenyiEq.sd(x,2)

data = rbinom(10,20,.5)
counts = tabulate(as.factor(data))
RenyiEq.sd(counts,2)
```

RenyiEq.z   *RenyiEq.z*

Description

Returns the Z estimator of Renyi's equivalent entropy. This estimator has exponentially decaying bias. When r=1 returns 1. See Zhang and Grabchak (2014a) for details.

Usage

```
RenyiEq.z(x, r)
```

Arguments

x	Vector of counts. Must be integer valued. Each entry represents the number of observations of a distinct letter.
r	Order of Renyi's equivalent entropy. Must be a strictly positive real number.

Author(s)

Lijuan Cao and Michael Grabchak

References

Z. Zhang and M. Grabchak (2014a). Entropic representation and estimation of diversity indices. http://arxiv.org/abs/1403.3031.

Examples

```
x = c(1,3,7,4,8)
RenyiEq.z(x,2)

data = rbinom(10,20,.5)
counts = tabulate(as.factor(data))
RenyiEq.z(counts,2)
```

SymKL.Plugin   *SymKL.Plugin*

Description

Returns the augmented plugin estimator of Symetrized Kullback-Leibler Divergence. See Zhang and Grabchak (2014b) for details.

Usage

```
SymKL.Plugin(x, y)
```

Arguments

x	Vector of counts from first distribution. Must be integer valued. Each entry represents the number of observations of a distinct letter.
y	Vector of counts from second distribution. Must be integer valued. Each entry represents the number of observations of a distinct letter.

Author(s)

Lijuan Cao and Michael Grabchak

References

Z. Zhang and M. Grabchak (2014b). Nonparametric Estimation of Kullback-Leibler Divergence. Neural Computation, DOI 10.1162/NECO_a_00646.

Examples

```
x = c(1,3,7,4,8) # first vector of counts
y = c(2,5,1,3,6) # second vector of counts
SymKL.Plugin(x,y)  # Estimated standard deviation
```

SymKL.sd   *SymKL.sd*

Description

Returns the estimated asymptotic standard deviation for the Z estimator of Symmetrized Kullback-Leibler's divergence. Note that this is also the asymptotic standard deviation of the plug-in estimator. See Zhang and Grabchak (2014b) for details.

Usage

```
SymKL.sd(x, y)
```

Arguments

x	Vector of counts from first distribution. Must be integer valued. Each entry represents the number of observations of a distinct letter.
y	Vector of counts from second distribution. Must be integer valued. Each entry represents the number of observations of a distinct letter.

Author(s)

Lijuan Cao and Michael Grabchak

References

Z. Zhang and M. Grabchak (2014b). Nonparametric Estimation of Kullback-Leibler Divergence. Neural Computation, DOI 10.1162/NECO_a_00646.

Examples

```
x = c(1,3,7,4,8) # first vector of counts
y = c(2,5,1,3,6) # second vector of counts
SymKL.sd(x,y)  # Estimated standard deviation
```

SymKL.z	*SymKL.z*

Description

Returns the Z estimator of Symetrized Kullback-Leibler Divergence, which has exponentialy decaying bias. See Zhang and Grabchak (2014b) for details.

Usage

```
SymKL.z(x, y)
```

Arguments

x	Vector of counts from first distribution. Must be integer valued. Each entry represents the number of observations of a distinct letter.
y	Vector of counts from second distribution. Must be integer valued. Each entry represents the number of observations of a distinct letter.

Author(s)

Lijuan Cao and Michael Grabchak

References

Z. Zhang and M. Grabchak (2014b). Nonparametric Estimation of Kullback-Leibler Divergence. Neural Computation, DOI 10.1162/NECO_a_00646.

Examples

```
x = c(1,3,7,4,8)
y = c(2,5,1,3,6)
SymKL.z(x,y)
```

Tsallis.sd	*Tsallis.sd*

Description

Returns the estimated asymptotic standard deviation for the Z estimator of Tsallis Entropy. Note that this is also the asymptotic standard deviation of the plug-in estimator. See Zhang and Grabchak (2014a) for details.

Usage

```
Tsallis.sd(x, r)
```

Arguments

x	Vector of counts. Must be integer valued. Each entry represents the number of observations of a distinct letter.
r	Order of Tsallis entropy. Must be a strictly positive real number. Not allowed to be 1, in that case use Entropy.sd instead.

Author(s)

Lijuan Cao and Michael Grabchak

References

Z. Zhang and M. Grabchak (2014a). Entropic representation and estimation of diversity indices. http://arxiv.org/abs/1403.3031.

Examples

```
x = c(1,3,7,4,8)
Tsallis.sd(x,2)

data = rbinom(10,20,.5)
counts = tabulate(as.factor(data))
Tsallis.sd(counts,2)
```

Tsallis.z   *Tsallis.z*

Description

Returns the Z estimator of Tsallis entropy. This is based on scaling and shifting the Z estimator of Renyi's equivalent entropy. When r=1 returns the Z estimator of Shannon's entropy. See Zhang and Grabchak (2014a) for details.

Usage

```
Tsallis.z(x, r)
```

Arguments

x	Vector of counts. Must be integer valued. Each entry represents the number of observations of a distinct letter.
r	Order or Renyi's equivalent entropy this index is based on. Must be a strictly positive real number.

Author(s)

Lijuan Cao and Michael Grabchak

References

Z. Zhang and M. Grabchak (2014a). Entropic representation and estimation of diversity indices. http://arxiv.org/abs/1403.3031.

Examples

```
x = c(1,3,7,4,8)
Tsallis.z(x,2)

data = rbinom(10,20,.5)
counts = tabulate(as.factor(data))
Tsallis.z(counts,2)
```

Index