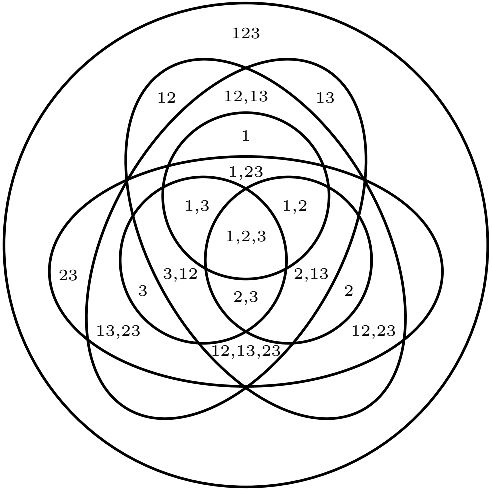
Information Decomposition

CSSS18 Tutorial

Conor Finn

July 6, 2018



Information Theory

- Entropy quantifies the average uncertainty of a variable

$$H(X) = \sum_{x \in X} p(x) \log {}^{1}\!/_{p(x)} \geq 0$$

Information Theory

- Entropy quantifies the average uncertainty of a variable

$$H(X) = \sum_{x \in X} p(x) \log {}^{1}/_{p(x)} \geq 0$$

- Entropy satisfies the Shannon inequalities

$$H(X),\ H(Y) \leq H(X,Y) \leq H(X) + H(Y)$$

Information Theory

- Entropy quantifies the average uncertainty of a variable

$$H(X) = \sum_{x \in X} p(x) \log {}^{1}/_{p(x)} \geq 0$$

- Entropy satisfies the Shannon inequalities

$$H(X),\ H(Y) \leq H(X,Y) \leq H(X) + H(Y)$$

$$H(X|Y) = H(X,Y) - H(Y) \geq 0$$
$$H(Y|X) = H(X,Y) - H(X) \geq 0$$
$$I(X;Y) \ = H(X) + H(Y) - H(X,Y) \geq 0$$

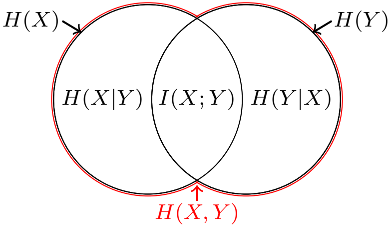
Information Theory

- Entropy quantifies the average uncertainty of a variable

$$H(X) = \sum_{x \in X} p(x) \log {}^{1}/_{p(x)} \geq 0$$

- Entropy satisfies the Shannon inequalities

$$H(X),\ H(Y) \leq H(X,Y) \leq H(X) + H(Y)$$



$$H(X|Y) = H(X,Y) - H(Y) \geq 0$$

$$H(Y|X) = H(X,Y) - H(X) \geq 0$$

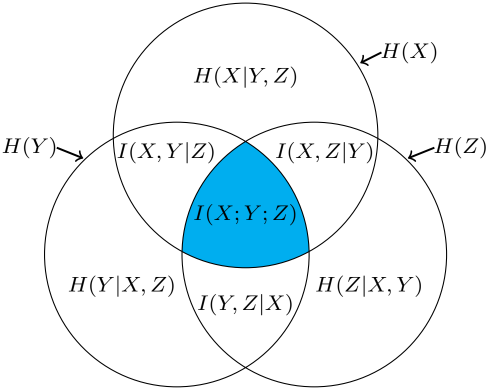
$$I(X;Y) = H(X) + H(Y) - H(X,Y) \geq 0$$

Multivariate Information Theory

- The mutual information for two variables

$$I(X;Y) = H(X) + H(Y) - H(X,Y) \geq 0.$$

Multivariate Information Theory



- The mutual information for two variables

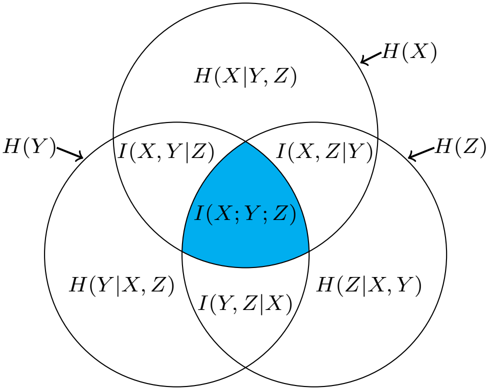
$$I(X;Y) = H(X) + H(Y) - H(X,Y) \geq 0.$$

- Multivariate mutual information

$$I(X;Y;X) = I(X;Y) + I(X;Z) - I(X;Y,Z).$$

a.k.a. interaction information, co-information

Multivariate Information Theory



- The mutual information for two variables

$$I(X;Y) = H(X) + H(Y) - H(X,Y) \geq 0.$$

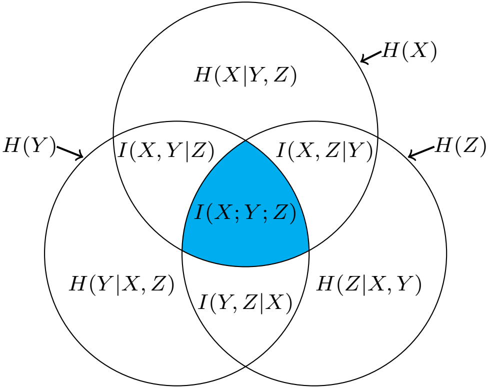
- Multivariate mutual information

$$I(X;Y;X) = I(X;Y) + I(X;Z) - I(X;Y,Z).$$

a.k.a. interaction information, co-information

- This quantity can be negative!
- How do we interpret this result?

Multivariate Information Theory



- The mutual information for two variables

$$I(X;Y) = H(X) + H(Y) - H(X,Y) \geq 0.$$

- Multivariate mutual information

$$I(X;Y;X) = I(X;Y) + I(X;Z) - I(X;Y,Z).$$

  a.k.a. interaction information, co-information

- This quantity can be negative!
- How do we interpret this result?

- There are no Shannon inequalities for multivariate information.

Unique, redundant, and synergistic information

Consider trying to predict T from S_1 and S_2

- Several types of information

Unique, redundant, and synergistic information

Consider trying to predict T from S_1 and S_2

- Several types of information
  - **Unique information** $U(S_1 \backslash S_2 \to T)$

Unq

p	s_1	s_2	t
1/4	0	0	0
1/4	0	1	0
1/4	1	0	1
1/4	1	1	1

Unique, redundant, and synergistic information

Consider trying to predict T from S_1 and S_2

- Several types of information
  - **Unique information** $U(S_1 \backslash S_2 \to T)$
  - **Redundant information** $R(S_1, S_2 \to T)$

UNQ

p	s_1	s_2	t
1/4	0	0	0
1/4	0	1	0
1/4	1	0	1
1/4	1	1	1

RDN

p	s_1	s_2	t
1/2	0	0	0
1/2	1	1	1

Unique, redundant, and synergistic information

Consider trying to predict T from S_1 and S_2

- Several types of information
  - **Unique information** $U(S_1 \backslash S_2 \to T)$
  - **Redundant information** $R(S_1, S_2 \to T)$
  - **Synergistic information** $C(S_1, S_2 \to T)$

Unq

p	s_1	s_2	t
1/4	0	0	0
1/4	0	1	0
1/4	1	0	1
1/4	1	1	1

Rdn

p	s_1	s_2	t
1/2	0	0	0
1/2	1	1	1

Xor

p	s_1	s_2	t
1/4	0	0	0
1/4	0	1	1
1/4	1	0	1
1/4	1	1	0

Unique, redundant, and synergistic information

Consider trying to predict T from S_1 and S_2

- Several types of information
  - **Unique information** $U(S_1 \backslash S_2 \to T)$
  - **Redundant information** $R(S_1, S_2 \to T)$
  - **Synergistic information** $C(S_1, S_2 \to T)$

Unq

p	s_1	s_2	t
1/4	0	0	0
1/4	0	1	0
1/4	1	0	1
1/4	1	1	1

Rdn

p	s_1	s_2	t
1/2	0	0	0
1/2	1	1	1

Xor

p	s_1	s_2	t
1/4	0	0	0
1/4	0	1	1
1/4	1	0	1
1/4	1	1	0

- Mutual information captures

$$I(T;S_1) = R(S_1, S_2 \to T) + U(S_1 \backslash S_2 \to T)$$
$$I(T;S_2) = R(S_1, S_2 \to T) + U(S_2 \backslash S_1 \to T)$$

- Joint mutual information captures

$$I(T;S_1,S_2) = R(S_1, S_2 \to T) + U(S_1 \backslash S_2 \to T) + U(S_2 \backslash S_1 \to T) + C(S_1, S_2 \to T)$$

Unique, redundant, and synergistic information

Consider trying to predict T from S_1 and S_2

- Several types of information
  - **Unique information** $U(S_1 \backslash S_2 \rightarrow T)$
  - **Redundant information** $R(S_1, S_2 \rightarrow T)$
  - **Synergistic information** $C(S_1, S_2 \rightarrow T)$

Unq

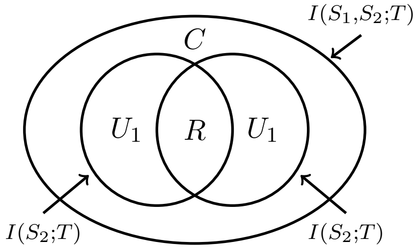
p	s_1	s_2	t
1/4	0	0	0
1/4	0	1	0
1/4	1	0	1
1/4	1	1	1

Rdn

p	s_1	s_2	t
1/2	0	0	0
1/2	1	1	1

Xor

p	s_1	s_2	t
1/4	0	0	0
1/4	0	1	1
1/4	1	0	1
1/4	1	1	0



- Mutual information captures

$$I(T;S_1) = R(S_1,S_2 \rightarrow T) + U(S_1 \backslash S_2 \rightarrow T)$$
$$I(T;S_2) = R(S_1,S_2 \rightarrow T) + U(S_2 \backslash S_1 \rightarrow T)$$

- Joint mutual information captures

$$I(T;S_1,S_2) = R(S_1,S_2 \rightarrow T) + U(S_1 \backslash S_2 \rightarrow T) + U(S_2 \backslash S_1 \rightarrow T) + C(S_1,S_2 \rightarrow T)$$

Different kinds of dependency

- Multivariate mutual information conflates redundant and synergistic information

$$
\begin{aligned}
I(T;S_1;S_2) &= I(T;S_1) + I(T;S_2) - I(S_1,S_2;T) \\
&= U(T:S_1 \setminus S_2) + R(T:S_1,S_2) + U(T:S_1 \setminus S_2) + R(T:S_1,S_2) \\
&\quad - R(S_1,S_2 \to T) + U(S_1 \; S_2 \to T) + U(S_2 \; S_1 \to T) + C(S_1,S_2 \to T) \\
&= R(S_1,S_2 \to T) - C(S_1,S_2 \to T)
\end{aligned}
$$

Different kinds of dependency

- Multivariate mutual information conflates redundant and synergistic information

$$
\begin{aligned}
I(T;S_1;S_2) &= I(T;S_1) + I(T;S_2) - I(S_1,S_2;T) \\
&= U(T:S_1 \backslash S_2) + R(T:S_1,S_2) + U(T:S_1 \backslash S_2) + R(T:S_1,S_2) \\
&\quad - R(S_1,S_2 \to T) + U(S_1\ S_2 \to T) + U(S_2\ S_1 \to T) + C(S_1,S_2 \to T) \\
&= R(S_1,S_2 \to T) - C(S_1,S_2 \to T)
\end{aligned}
$$

- How can we separate these effects in general?

Partial information decomposition

Framework from Williams and Beer (2010)

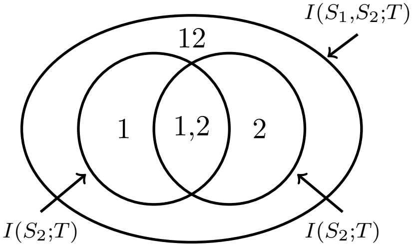
- Axioms for redundant information
  1. Commutativity
  2. Monotonically decreasing
  3. Self-redundancy (idempotency)
- Yields a redundancy lattice

Partial information decomposition

Framework from Williams and Beer (2010)

- Axioms for redundant information
  1. Commutativity
  2. Monotonically decreasing
  3. Self-redundancy (idempotency)

- Yields a redundancy lattice

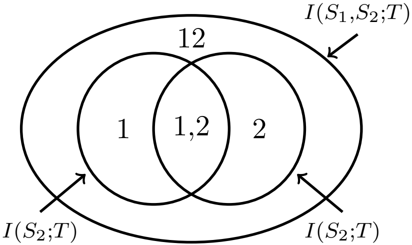


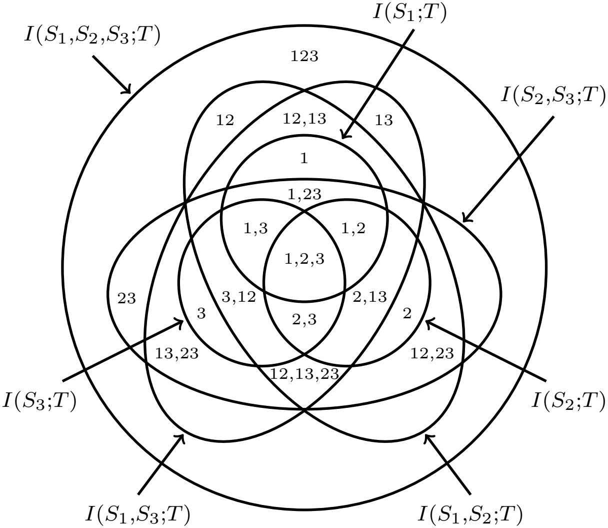
Partial information decomposition

Framework from Williams and Beer (2010)

- Axioms for redundant information
  1. Commutativity
  2. Monotonically decreasing
  3. Self-redundancy (idempotency)

- Yields a redundancy lattice





PID is elegant, however...

PID is elegant, however...

- Defining the redundant information is not simple

PID is elegant, however...

- Defining the redundant information is not simple
- Why should we be interested?
  - Feature selection in machine learning
  - Network information theory
  - Lossless ompression of structured databases

PID is elegant, however...

- Defining the redundant information is not simple
- Why should we be interested?
  - Feature selection in machine learning
  - Network information theory
  - Lossless ompression of structured databases
  - Neuroscience: quantifying information modification in the brain

PID is elegant, however...

- Defining the redundant information is not simple
- Why should we be interested?
  - Feature selection in machine learning
  - Network information theory
  - Lossless ompression of structured databases
  - Neuroscience: quantifying information modification in the brain
- The good news: pointwise partial information decomposition


entropy

MDPI

Article

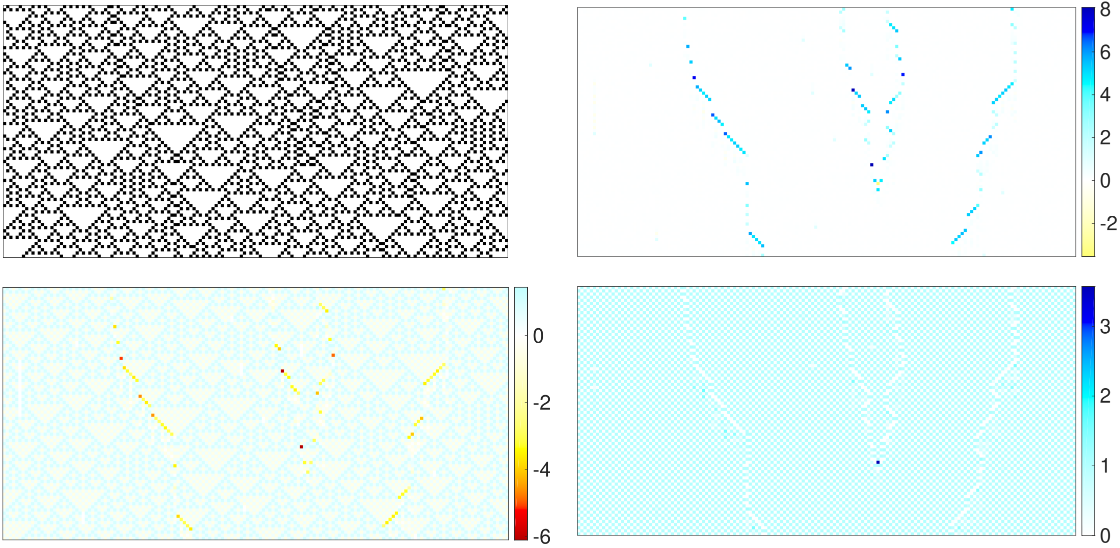
Pointwise Partial Information Decomposition Using the Specificity and Ambiguity Lattices

Conor Finn [1,2,*] and Joseph T. Lizier [1]

1 Complex Systems Research Group and Centre for Complex Systems, Faculty of Engineering & IT, The University of Sydney, NSW 2006, Australia; joseph.lizier@sydney.edu.au
2 CSIRO Data61, Marsfield NSW 2122, Australia
* Correspondence: conor.finn@sydney.edu.au

Received: 10 July 2017; Accepted: 10 April 2018; Published: 18 April 2018

Quantifying information modification in CAs



Questions?

Redundancy measures: $\mathrm{I}_{\mathsf{min}}$

Original measure of redundancy introduced by Williams and Beer

$$\mathrm{I}_{\mathsf{min}}(X : Y_1, \ldots, Y_k) = \sum_x p(x) \min_{Y_i} \mathrm{I}(X = x; Y_i)$$

Redundancy measures: I_{min}

Original measure of redundancy introduced by Williams and Beer

$$\mathrm{I}_{\text{min}}(X : Y_1, \ldots, Y_k) = \sum_x p(x) \min_{Y_i} \mathrm{I}(X = x; Y_i)$$

- Semi-local approach: for each $X = x$ the redundant information is the minimum information provided by all of the sources Y_i

Redundancy measures: I_{min}

Original measure of redundancy introduced by Williams and Beer

$$\mathrm{I}_{\text{min}}(X : Y_1, \ldots, Y_k) = \sum_x p(x) \min_{Y_i} \mathrm{I}(X = x; Y_i)$$

- Semi-local approach: for each $X = x$ the redundant information is the minimum information provided by all of the sources Y_i
- Widely critised after its introduction — two bit copy problem

X	Y	Z	P
00	0	0	1/4
01	0	1	1/4
10	1	0	1/4
01	1	1	1/4

Redundancy measures: $\mathrm{I}_{\mathsf{min}}$

Original measure of redundancy introduced by Williams and Beer

$$\mathrm{I}_{\mathsf{min}}(X : Y_1, \ldots, Y_k) = \sum_x p(x) \min_{Y_i} \mathrm{I}(X = x; Y_i)$$

- Semi-local approach: for each $X = x$ the redundant information is the minimum information provided by all of the sources Y_i
- Widely critised after its introduction — two bit copy problem

X	Y	Z	P
00	0	0	1/4
01	0	1	1/4
10	1	0	1/4
01	1	1	1/4

$\mathrm{I}_{\mathsf{min}}(X : Y; Z) = 1$ bit

Redundancy measures: I_{min}

Original measure of redundancy introduced by Williams and Beer

$$\mathrm{I}_{\text{min}}(X : Y_1, \ldots, Y_k) = \sum_x p(x) \min_{Y_i} \mathrm{I}(X = x; Y_i)$$

- Semi-local approach: for each $X = x$ the redundant information is the minimum information provided by all of the sources Y_i
- Widely critised after its introduction — two bit copy problem

X	Y	Z	P
00	0	0	1/4
01	0	1	1/4
10	1	0	1/4
01	1	1	1/4

$$\mathrm{I}_{\text{min}}(X : Y; Z) = 1 \text{ bit}$$

  - “The problem is I_{min} does not distinguish whether sources carry *the same* information or just *the same amount* of information”

Redundancy measures: I_{red}

Based on information geometry and introduced by Harder et al.

$$\mathrm{I}_{\text{red}}(Z : X; Y) = \min\left\{ \mathrm{I}_Z^{\pi}(X \searrow Y)\,,\ \mathrm{I}_Z^{\pi}(X \searrow Y) \right\}$$

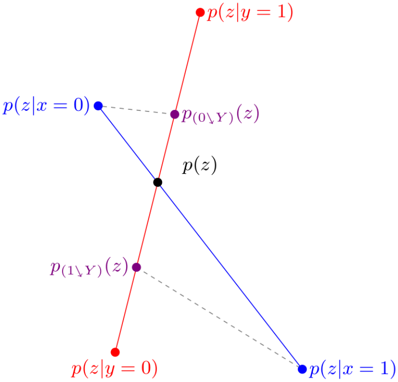
where $\mathrm{I}_Z^{\pi}(X \searrow Y)$ is the mutual information between Z and X expressed in terms of the mutual information between Z and Y.

Redundancy measures: $\mathrm{I}_{\mathsf{red}}$

Based on information geometry and introduced by Harder et al.

$$\mathrm{I}_{\mathsf{red}}(Z:X;Y) = \min\left\{\, \mathrm{I}_Z^{\pi}(X \searrow Y)\,,\ \mathrm{I}_Z^{\pi}(X \searrow Y)\right\}$$

where $\mathrm{I}_Z^{\pi}(X \searrow Y)$ is the mutual information between Z and X expressed in terms of the mutual information between Z and Y.

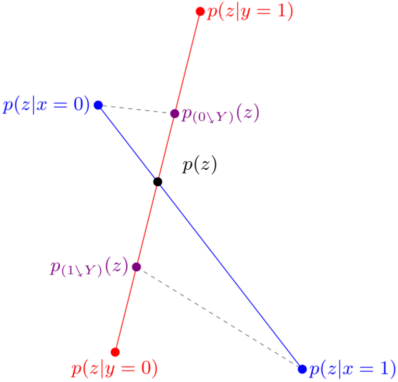


Redundancy measures: I_{red}

Based on information geometry and introduced by Harder et al.

$$\mathrm{I}_{\text{red}}(Z : X; Y) = \min \left\{ \mathrm{I}_Z^{\pi}(X \searrow Y) \,,\, \mathrm{I}_Z^{\pi}(X \searrow Y) \right\}$$

where $\mathrm{I}_Z^{\pi}(X \searrow Y)$ is the mutual information between Z and X expressed in terms of the mutual information between Z and Y.



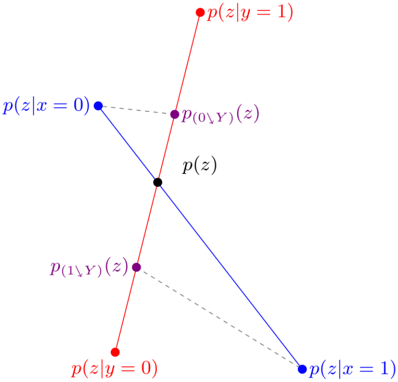
- Only able to quantify bivariate redundancy: multivariate extension highly non-trivial and evaluation is intractable

Redundancy measures: I_{red}

Based on information geometry and introduced by Harder et al.

$$\mathrm{I}_{\text{red}}(Z : X; Y) = \min\left\{ \mathrm{I}_Z^{\pi}(X \searrow Y)\,,\ \mathrm{I}_Z^{\pi}(X \searrow Y) \right\}$$

where $\mathrm{I}_Z^{\pi}(X \searrow Y)$ is the mutual information between Z and X expressed in terms of the mutual information between Z and Y.



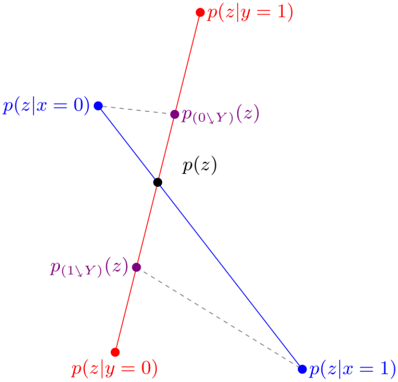
- Only able to quantify bivariate redundancy: multivariate extension highly non-trivial and evaluation is intractable
- Not even clear that it does indeed capture the redundant information

Redundancy measures: $\mathrm{I}_{\mathsf{red}}$

Based on information geometry and introduced by Harder et al.

$$\mathrm{I}_{\mathsf{red}}(Z : X; Y) = \min \left\{ \mathrm{I}_Z^{\pi}(X \searrow Y) \, , \, \mathrm{I}_Z^{\pi}(X \searrow Y) \right\}$$

where $\mathrm{I}_Z^{\pi}(X \searrow Y)$ is the mutual information between Z and X expressed in terms of the mutual information between Z and Y.



- Only able to quantify bivariate redundancy: multivariate extension highly non-trivial and evaluation is intractable
- Not even clear that it does indeed capture the redundant information
- No meaningful local intepretation

Unique information measure: $\widetilde{\mathrm{UI}}$

Indroduced by Bertschinger et al. — game-theoretic motivation

- Defining the unique information implicitly defines the redundant information in the partial information decomposition framework

Unique information measure: $\widetilde{\mathrm{UI}}$

Indroduced by Bertschinger et al. — game-theoretic motivation

- Defining the unique information implicitly defines the redundant information in the partial information decomposition framework
- If a source contains unique information then there must be a way to exploit this information in a decision problem

Unique information measure: $\widetilde{\mathrm{UI}}$

Indroduced by Bertschinger et al. — game-theoretic motivation

- Defining the unique information implicitly defines the redundant information in the partial information decomposition framework
- If a source contains unique information then there must be a way to exploit this information in a decision problem
- No unique local intepretation

Unique information measure: $\widetilde{\mathrm{UI}}$

Indroduced by Bertschinger et al. — game-theoretic motivation

- Defining the unique information implicitly defines the redundant information in the partial information decomposition framework
- If a source contains unique information then there must be a way to exploit this information in a decision problem
- No unique local intepretation
- Worse than that

X	Y	Z	P
0	0	0	1/2
1	0	1	1/4
1	1	0	1/4

Unique information measure: $\widetilde{\mathrm{UI}}$

Indroduced by Bertschinger et al. — game-theoretic motivation

- Defining the unique information implicitly defines the redundant information in the partial information decomposition framework
- If a source contains unique information then there must be a way to exploit this information in a decision problem
- No unique local intepretation
- Worse than that

X	Y	Z	P
0	0	0	1/2
1	0	1	1/4
1	1	0	1/4

$$\widetilde{\mathrm{UI}}(X : Y) = \widetilde{\mathrm{UI}}(X : Y) = 0 \text{ bit}$$

References

Paul L Williams and Randall D Beer. Nonnegative decomposition of multivariate information. *arXiv preprint arXiv:1004.2515*, 2010.