Redundant information, total information and higher order interactions

Conor Finn

DEMICS – MPI PKS

June 6, 2023

Max Planck Institute for Mathematics in the Sciences

MAX-PLANCK-GESELLSCHAFT

Overview

1. The functional form of redundant information

2. The total information lattice

Section 1: The functional form of redundant information

Functional dependence of redundant information

- Does the redundant information depend on the full distribution,

$$I_\cap(T; \boldsymbol{A}_1, \ldots, \boldsymbol{A}_k) = f\big(P(T, \boldsymbol{A}_1, \ldots, \boldsymbol{A}_k)\big)?$$

- Or does it only depend on the marginal distributions,

$$I_\cap(T; \boldsymbol{A}_1, \ldots, \boldsymbol{A}_k) = f\big(P(T, \boldsymbol{A}_1), \ldots, P(T, \boldsymbol{A}_k)\big)?$$

Sources

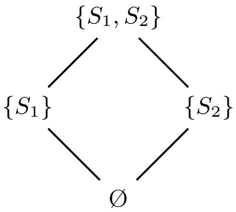
- Consider a set of variables $\boldsymbol{S} = \{S_1, \ldots, S_n\}$ and let $\boldsymbol{A}$ be a subset called a source

Sources

- Consider a set of variables $\boldsymbol{S} = \{S_1, \ldots, S_n\}$ and let $\boldsymbol{A}$ be a subset called a source
- The set of all sources corresponds to the power set of $\boldsymbol{S}$
- Power set can be ordered by set inclusion yielding the inclusion lattice

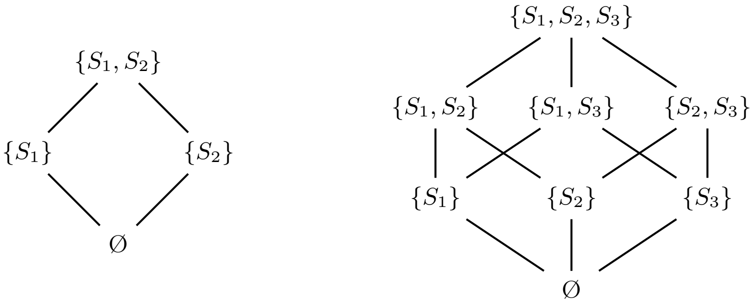
Sources

- Consider a set of variables $\boldsymbol{S} = \{S_1, \ldots, S_n\}$ and let $\boldsymbol{A}$ be a subset called a source
- The set of all sources corresponds to the power set of $\boldsymbol{S}$
- Power set can be ordered by set inclusion yielding the inclusion lattice



Sources

- Consider a set of variables $\boldsymbol{S} = \{S_1, \ldots, S_n\}$ and let $\boldsymbol{A}$ be a subset called a source
- The set of all sources corresponds to the power set of $\boldsymbol{S}$
- Power set can be ordered by set inclusion yielding the inclusion lattice



Partial information decomposition

- Mutual information $I(T; \boldsymbol{A})$ quantifies the information provided by a single source
- Define a function $I_\cap$ that quantifies the redundant info provided by multiple sources

Partial information decomposition

- Mutual information $I(T; \boldsymbol{A})$ quantifies the information provided by a single source
- Define a function $I_\cap$ that quantifies the redundant info provided by multiple sources

The Williams and Beer Axioms

1. *Symmetry*: $I_\cap(T; \boldsymbol{A}_1, \ldots, \boldsymbol{A}_k)$ is invariant under permutations of the $\boldsymbol{A}_i$'s
2. *Self-redundancy*: $I_\cap(T; \boldsymbol{A}_i) = I(T; \boldsymbol{A})$
3. *Monotonicity*:

$$I_\cap(T; \boldsymbol{A}_1, \ldots, \boldsymbol{A}_k) \leq I_\cap(T; \boldsymbol{A}_1, \ldots, \boldsymbol{A}_{k-1})$$

with equality if $\boldsymbol{A}_{k-1} \subseteq \boldsymbol{A}_k$

Partial information decomposition

- Mutual information $I(T; \boldsymbol{A})$ quantifies the information provided by a single source
- Define a function $I_\cap$ that quantifies the redundant info provided by multiple sources

The Williams and Beer Axioms

1. *Symmetry*: $I_\cap(T; \boldsymbol{A}_1, \ldots, \boldsymbol{A}_k)$ is invariant under permutations of the $\boldsymbol{A}_i$'s
2. *Self-redundancy*: $I_\cap(T; \boldsymbol{A}_i) = I(T; \boldsymbol{A})$
3. *Monotonicity*:

$$I_\cap(T; \boldsymbol{A}_1, \ldots, \boldsymbol{A}_k) \leq I_\cap(T; \boldsymbol{A}_1, \ldots, \boldsymbol{A}_{k-1})$$

with equality if $\boldsymbol{A}_{k-1} \subseteq \boldsymbol{A}_k$

- What are the different ways the sources can provide redundant information?
  - Answering this question corresponds to determining the domain of $I_\cap$

Domain of the redundancy function

- $I_\cap$ can be applied to any collection of sources, e.g. $\{\{S_1\},\{S_2,S_3\}\}$

Domain of the redundancy function

- $I_\cap$ can be applied to any collection of sources, e.g. $\{\{S_1\}, \{S_2, S_3\}\}$
- Many collection are equivalent due to Axiom 3,

Domain of the redundancy function

- $I_\cap$ can be applied to any collection of sources, e.g. $\{\{S_1\},\{S_2,S_3\}\}$
- Many collection are equivalent due to Axiom 3, e.g. since $\{S_1\} \subseteq \{S_1,S_2\}$,

$$I_\cap(T;\{S_1\},\{S_1,S_2\}) = I_\cap(T;\{S_1\})$$

Domain of the redundancy function

- $I_\cap$ can be applied to any collection of sources, e.g. $\{\{S_1\}, \{S_2, S_3\}\}$
- Many collection are equivalent due to Axiom 3, e.g. since $\{S_1\} \subseteq \{S_1, S_2\}$,

$$I_\cap(T; \{S_1\}, \{S_1, S_2\}) = I_\cap(T; \{S_1\})$$

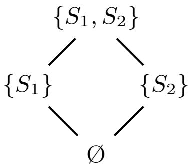
- The unique collection are those such that no source is a subset of any other
  - i.e. the antichains of the inclusion lattice

Domain of the redundancy function

- $I_\cap$ can be applied to any collection of sources, e.g. $\{\{S_1\},\{S_2,S_3\}\}$
- Many collection are equivalent due to Axiom 3, e.g. since $\{S_1\} \subseteq \{S_1, S_2\}$,

$$I_\cap(T;\{S_1\},\{S_1,S_2\}) = I_\cap(T;\{S_1\})$$

- The unique collection are those such that no source is a subset of any other
  - i.e. the antichains of the inclusion lattice



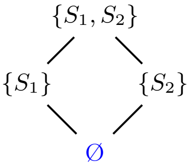
- Antichains:

Domain of the redundancy function

- $I_\cap$ can be applied to any collection of sources, e.g. $\{\{S_1\}, \{S_2, S_3\}\}$
- Many collection are equivalent due to Axiom 3, e.g. since $\{S_1\} \subseteq \{S_1, S_2\}$,

$$I_\cap(T; \{S_1\}, \{S_1, S_2\}) = I_\cap(T; \{S_1\})$$

- The unique collection are those such that no source is a subset of any other
  - i.e. the antichains of the inclusion lattice



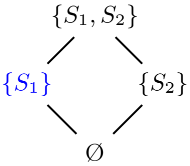
- Antichains: $\{\emptyset\}$

Domain of the redundancy function

- $I_\cap$ can be applied to any collection of sources, e.g. $\{\{S_1\},\{S_2,S_3\}\}$
- Many collection are equivalent due to Axiom 3, e.g. since $\{S_1\} \subseteq \{S_1,S_2\}$,

$$I_\cap(T;\{S_1\},\{S_1,S_2\}) = I_\cap(T;\{S_1\})$$

- The unique collection are those such that no source is a subset of any other
  - i.e. the antichains of the inclusion lattice



- Antichains: $\{\emptyset\}$ $\{\{S_1\}\}$,$\{\{S_1\}\}$,

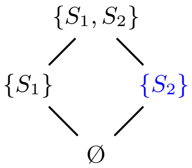
Domain of the redundancy function

- $I_\cap$ can be applied to any collection of sources, e.g. $\{\{S_1\}, \{S_2, S_3\}\}$

- Many collection are equivalent due to Axiom 3, e.g. since $\{S_1\} \subseteq \{S_1, S_2\}$,

$$I_\cap(T; \{S_1\}, \{S_1, S_2\}) = I_\cap(T; \{S_1\})$$

- The unique collection are those such that no source is a subset of any other
  - i.e. the antichains of the inclusion lattice



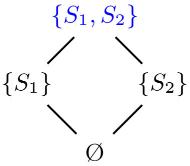
- Antichains: $\{\emptyset\}$ $\{\{S_1\}\}$,$\{\{S_1\}\}$, $\{\{S_2\}\}$,

Domain of the redundancy function

- $I_\cap$ can be applied to any collection of sources, e.g. $\{\{S_1\}, \{S_2, S_3\}\}$
- Many collection are equivalent due to Axiom 3, e.g. since $\{S_1\} \subseteq \{S_1, S_2\}$,

$$I_\cap(T; \{S_1\}, \{S_1, S_2\}) = I_\cap(T; \{S_1\})$$

- The unique collection are those such that no source is a subset of any other
  - i.e. the antichains of the inclusion lattice



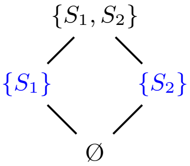
- Antichains: $\{\emptyset\}$ $\{\{S_1\}\}$,$\{\{S_1\}\}$, $\{\{S_2\}\}$, $\{\{S_1, S_2\}\}$ and

Domain of the redundancy function

- $I_\cap$ can be applied to any collection of sources, e.g. $\{\{S_1\}, \{S_2, S_3\}\}$
- Many collection are equivalent due to Axiom 3, e.g. since $\{S_1\} \subseteq \{S_1, S_2\}$,

$$I_\cap(T; \{S_1\}, \{S_1, S_2\}) = I_\cap(T; \{S_1\})$$

- The unique collection are those such that no source is a subset of any other
  - i.e. the antichains of the inclusion lattice



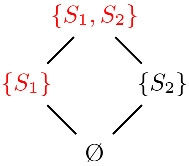
- Antichains: $\{\emptyset\}$ $\{\{S_1\}\}$,$\{\{S_1\}\}$, $\{\{S_2\}\}$, $\{\{S_1, S_2\}\}$ and $\{\{S_1\}, \{S_2\}\}$

Domain of the redundancy function

- $I_\cap$ can be applied to any collection of sources, e.g. $\{\{S_1\},\{S_2,S_3\}\}$
- Many collection are equivalent due to Axiom 3, e.g. since $\{S_1\} \subseteq \{S_1,S_2\}$,

$$I_\cap(T;\{S_1\},\{S_1,S_2\}) = I_\cap(T;\{S_1\})$$

- The unique collection are those such that no source is a subset of any other
  - i.e. the antichains of the inclusion lattice



- Antichains: $\{\emptyset\}$ $\{\{S_1\}\}$,$\{\{S_1\}\}$, $\{\{S_2\}\}$, $\{\{S_1,S_2\}\}$ and $\{\{S_1\},\{S_2\}\}$

Redundancy lattice for sets of variables

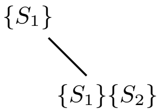
- Axiom 3 also defines a partial order $\preccurlyeq_{\cap}$ over these antichains

Redundancy lattice for sets of variables

- Axiom 3 also defines a partial order $\preccurlyeq_\cap$ over these antichains
  - $I_\cap(T;\{S_1\},\{S_2\}) \leq I_\cap(T;\{S_1\})$ follows directly

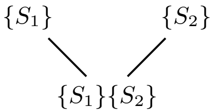
Redundancy lattice for sets of variables

- Axiom 3 also defines a partial order $\preccurlyeq_\cap$ over these antichains
  - $I_\cap(T;\{S_1\},\{S_2\}) \leq I_\cap(T;\{S_1\})$ follows directly



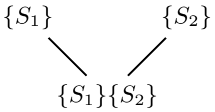
Redundancy lattice for sets of variables

- Axiom 3 also defines a partial order $\preccurlyeq_\cap$ over these antichains
  - $I_\cap(T; \{S_1\}, \{S_2\}) \leq I_\cap(T; \{S_1\})$ follows directly



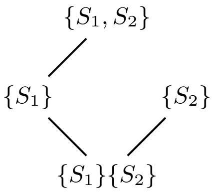
Redundancy lattice for sets of variables

- Axiom 3 also defines a partial order $\preccurlyeq_\cap$ over these antichains
  - $I_\cap(T;\{S_1\},\{S_2\}) \leq I_\cap(T;\{S_1\})$ follows directly
  - $I_\cap(T;\{S_1\}) = I_\cap(T;\{S_1\},\{S_1,S_2\}) \leq I_\cap(T;\{S_1,S_2\})$ since $\{S_1\} \subseteq \{S_1,S_2\}$



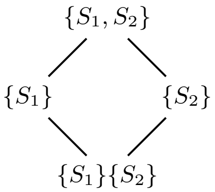
Redundancy lattice for sets of variables

- Axiom 3 also defines a partial order $\preccurlyeq_\cap$ over these antichains
  - $I_\cap(T;\{S_1\},\{S_2\}) \leq I_\cap(T;\{S_1\})$ follows directly
  - $I_\cap(T;\{S_1\}) = I_\cap(T;\{S_1\},\{S_1,S_2\}) \leq I_\cap(T;\{S_1,S_2\})$ since $\{S_1\} \subseteq \{S_1,S_2\}$



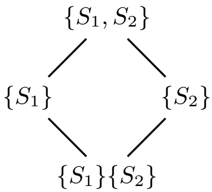
Redundancy lattice for sets of variables

- Axiom 3 also defines a partial order $\preccurlyeq_\cap$ over these antichains
  - $I_\cap(T; \{S_1\}, \{S_2\}) \leq I_\cap(T; \{S_1\})$ follows directly
  - $I_\cap(T; \{S_1\}) = I_\cap(T; \{S_1\}, \{S_1, S_2\}) \leq I_\cap(T; \{S_1, S_2\})$ since $\{S_1\} \subseteq \{S_1, S_2\}$



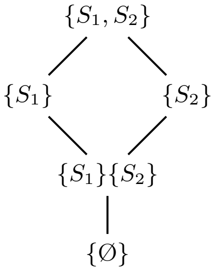
Redundancy lattice for sets of variables

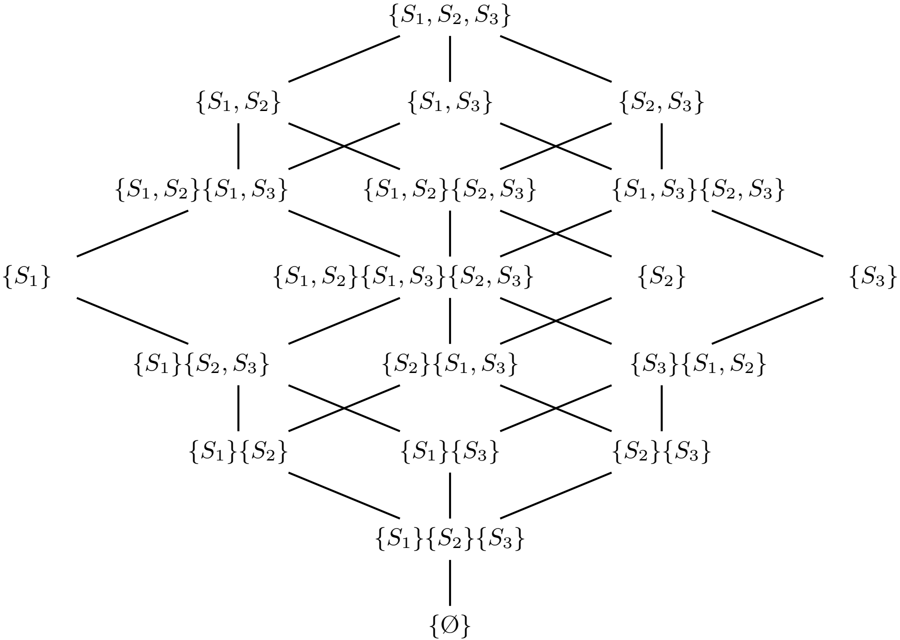
- Axiom 3 also defines a partial order $\preccurlyeq_\cap$ over these antichains
  - $I_\cap(T; \{S_1\}, \{S_2\}) \leq I_\cap(T; \{S_1\})$ follows directly
  - $I_\cap(T; \{S_1\}) = I_\cap(T; \{S_1\}, \{S_1, S_2\}) \leq I_\cap(T; \{S_1, S_2\})$ since $\{S_1\} \subseteq \{S_1, S_2\}$
  - $I_\cap(T; \emptyset) \leq I_\cap(T; \{S_1\}, \{S_1, S_2\})$ since $\{\emptyset\} \subseteq \{S_1\}$ and $\{\emptyset\} \subseteq \{S_2\}$



Redundancy lattice for sets of variables

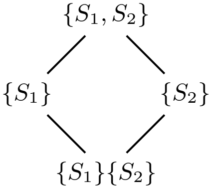
- Axiom 3 also defines a partial order $\preccurlyeq_\cap$ over these antichains
  - $I_\cap(T;\{S_1\},\{S_2\}) \leq I_\cap(T;\{S_1\})$ follows directly
  - $I_\cap(T;\{S_1\}) = I_\cap(T;\{S_1\},\{S_1,S_2\}) \leq I_\cap(T;\{S_1,S_2\})$ since $\{S_1\} \subseteq \{S_1,S_2\}$
  - $I_\cap(T;\emptyset) \leq I_\cap(T;\{S_1\},\{S_1,S_2\})$ since $\{\emptyset\} \subseteq \{S_1\}$ and $\{\emptyset\} \subseteq \{S_2\}$




{S1, S2, S3}
{S1, S2}
{S1, S3}
{S2, S3}
{S1, S2}{S1, S3}
{S1, S2}{S2, S3}
{S1, S3}{S2, S3}
{S1}
{S1, S2}{S1, S3}{S2, S3}
{S2}
{S3}
{S1}{S2, S3}
{S2}{S1, S3}
{S3}{S1, S2}
{S1}{S2}
{S1}{S3}
{S2}{S3}
{S1}{S2}{S3}
{∅}

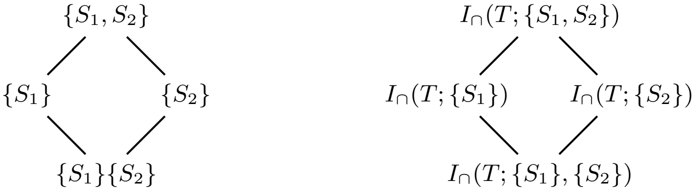
Missing details

- Target variable T does not appear in the redundancy lattice



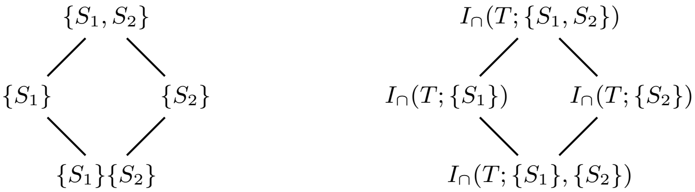
Missing details

- Target variable T does not appear in the redundancy lattice



Missing details

- Target variable T does not appear in the redundancy lattice



- Leaves the door open to differing interpretations:
  - Redundant information could depend on the full distribution

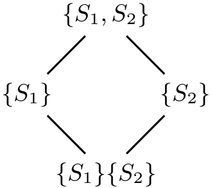
$$I_\cap(T; \{S_1\}, \{S_2\}) = f\big(P(T, S_1, S_2)\big)$$

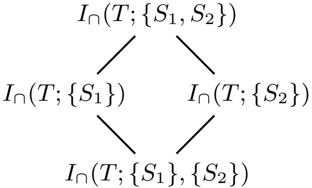
  - Or it could depend only on the marginal distributions

$$I_\cap(T; \{S_1\}, \{S_2\}) = f\big(P(T, S_1), P(T, S_2)\big)$$

Missing details

- Target variable T does not appear in the redundancy lattice





- Leaves the door open to differing interpretations:
  - Redundant information could depend on the full distribution

$$I_\cap(T; \{S_1\}, \{S_2\}) = f\big(P(T, S_1, S_2)\big)$$

  - Or it could depend only on the marginal distributions

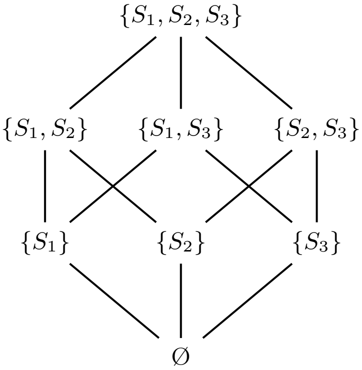
$$I_\cap(T; \{S_1\}, \{S_2\}) = f\big(P(T, S_1), P(T, S_2)\big)$$

- Why are we talking about sets of random variables?

Lattice of marginal random vectors

- Variables should be from the same probability space

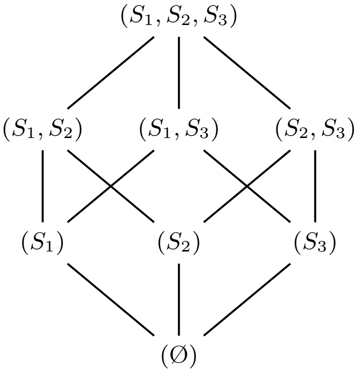
$$\boldsymbol{S} = \{S_1, S_2, S_3\} \to (\boldsymbol{S}) = (S_1, S_2, S_3)$$



Lattice of marginal random vectors

- Variables should be from the same probability space

$$\boldsymbol{S} = \{S_1, S_2, S_3\} \to (\boldsymbol{S}) = (S_1, S_2, S_3)$$



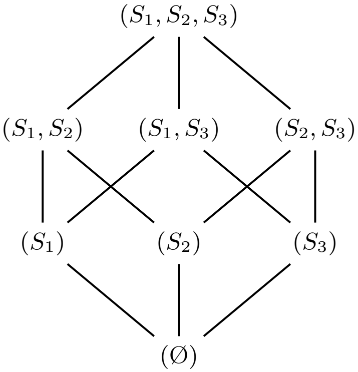
Lattice of marginal random vectors

- Variables should be from the same probability space

$$\boldsymbol{S} = \{S_1, S_2, S_3\} \rightarrow (\boldsymbol{S}) = (S_1, S_2, S_3)$$

- Source $\boldsymbol{A}$ corresponds to a marginal vector of $\boldsymbol{S}$, e.g.

$$\boldsymbol{A} = \{S_1, S_2\} \rightarrow (\boldsymbol{A}) = (S_1, S_2)$$



Lattice of marginal random vectors

- Variables should be from the same probability space

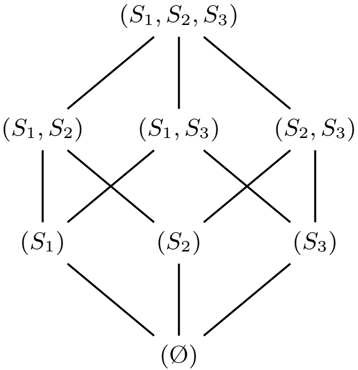
$$\boldsymbol{S} = \{S_1, S_2, S_3\} \to (\boldsymbol{S}) = (S_1, S_2, S_3)$$

- Source $\boldsymbol{A}$ corresponds to a marginal vector of $\boldsymbol{S}$, e.g.

$$\boldsymbol{A} = \{S_1, S_2\} \to (\boldsymbol{A}) = (S_1, S_2)$$

- Mutual information: $I(T; \boldsymbol{A}) = f((T, \boldsymbol{A}))$

$$(\boldsymbol{S}) = (S_1, S_2, S_3) \to (T, \boldsymbol{S}) = (T, S_1, S_2, S_3)$$



Lattice of marginal random vectors

- Variables should be from the same probability space

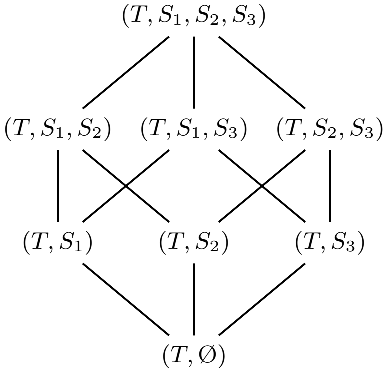
$$\boldsymbol{S} = \{S_1, S_2, S_3\} \to (\boldsymbol{S}) = (S_1, S_2, S_3)$$

- Source $\boldsymbol{A}$ corresponds to a marginal vector of $\boldsymbol{S}$, e.g.

$$\boldsymbol{A} = \{S_1, S_2\} \to (\boldsymbol{A}) = (S_1, S_2)$$

- Mutual information: $I(T; \boldsymbol{A}) = f((T, \boldsymbol{A}))$

$$(\boldsymbol{S}) = (S_1, S_2, S_3) \to (T, \boldsymbol{S}) = (T, S_1, S_2, S_3)$$



Lattice of marginal random vectors

- Variables should be from the same probability space

$$\boldsymbol{S} = \{S_1, S_2, S_3\} \to (\boldsymbol{S}) = (S_1, S_2, S_3)$$

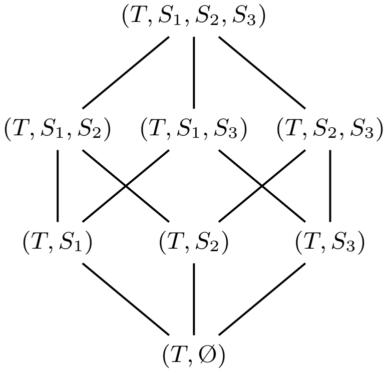
- Source $\boldsymbol{A}$ corresponds to a marginal vector of $\boldsymbol{S}$, e.g.

$$\boldsymbol{A} = \{S_1, S_2\} \to (\boldsymbol{A}) = (S_1, S_2)$$

- Mutual information: $I(T; \boldsymbol{A}) = f((T, \boldsymbol{A}))$

$$(\boldsymbol{S}) = (S_1, S_2, S_3) \to (T, \boldsymbol{S}) = (T, S_1, S_2, S_3)$$

- Lattice of all marginal vectors of $(T, \boldsymbol{S})$ containing T



Lattice of marginal random vectors

- Variables should be from the same probability space

$$\boldsymbol{S} = \{S_1, S_2, S_3\} \to (\boldsymbol{S}) = (S_1, S_2, S_3)$$

- Source $\boldsymbol{A}$ corresponds to a marginal vector of $\boldsymbol{S}$, e.g.

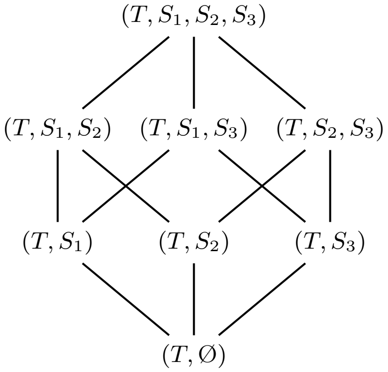
$$\boldsymbol{A} = \{S_1, S_2\} \to (\boldsymbol{A}) = (S_1, S_2)$$

- Mutual information: $I(T; \boldsymbol{A}) = f\big((T, \boldsymbol{A})\big)$

$$(\boldsymbol{S}) = (S_1, S_2, S_3) \to (T, \boldsymbol{S}) = (T, S_1, S_2, S_3)$$

- Lattice of all marginal vectors of $(T, \boldsymbol{S})$ containing T

- Mutual information is a monotonic, bottom normalised lattice function of this lattice



Lattice of marginal random vectors

- Variables should be from the same probability space

$$\boldsymbol{S} = \{S_1, S_2, S_3\} \to (\boldsymbol{S}) = (S_1, S_2, S_3)$$

- Source $\boldsymbol{A}$ corresponds to a marginal vector of $\boldsymbol{S}$, e.g.

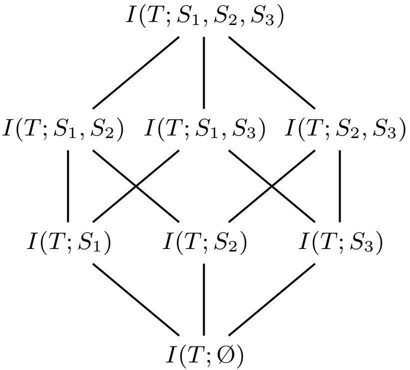
$$\boldsymbol{A} = \{S_1, S_2\} \to (\boldsymbol{A}) = (S_1, S_2)$$

- Mutual information: $I(T; \boldsymbol{A}) = f\big((T, \boldsymbol{A})\big)$

$$(\boldsymbol{S}) = (S_1, S_2, S_3) \to (T, \boldsymbol{S}) = (T, S_1, S_2, S_3)$$

- Lattice of all marginal vectors of $(T, \boldsymbol{S})$ containing T

- Mutual information is a monotonic, bottom normalised lattice function of this lattice



Lattice of marginal random vectors

- Variables should be from the same probability space

$$\boldsymbol{S} = \{S_1, S_2, S_3\} \to (\boldsymbol{S}) = (S_1, S_2, S_3)$$

- Source $\boldsymbol{A}$ corresponds to a marginal vector of $\boldsymbol{S}$, e.g.

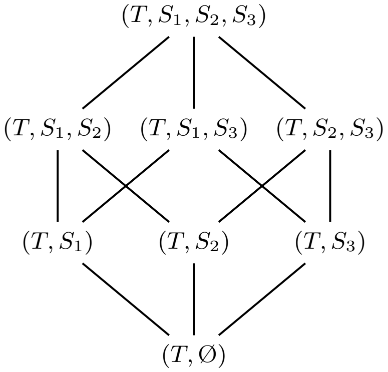
$$\boldsymbol{A} = \{S_1, S_2\} \to (\boldsymbol{A}) = (S_1, S_2)$$

- Mutual information: $I(T; \boldsymbol{A}) = f\big((T, \boldsymbol{A})\big)$

$$(\boldsymbol{S}) = (S_1, S_2, S_3) \to (T, \boldsymbol{S}) = (T, S_1, S_2, S_3)$$

- Lattice of all marginal vectors of $(T, \boldsymbol{S})$ containing T

- Mutual information is a monotonic, bottom normalised lattice function of this lattice



Domain of the redundancy function for vectors

- Axiom 2 specifies the function form of redundant information for a single source

$$I_{\cap}(T;\boldsymbol{A}) = I(T;\boldsymbol{A}) = f\big((T,\boldsymbol{A})\big)$$

Domain of the redundancy function for vectors

- Axiom 2 specifies the function form of redundant information for a single source

$$I_\cap(T; \boldsymbol{A}) = I(T; \boldsymbol{A}) = f\big((T, \boldsymbol{A})\big)$$

- Extending this dependence to multiple marginal vectors

$$I_\cap(T; \boldsymbol{A}_1, \ldots, \boldsymbol{A}_k) = f\big((T, \boldsymbol{A}_1), \ldots, (T, \boldsymbol{A}_k)\big)$$

Domain of the redundancy function for vectors

- Axiom 2 specifies the function form of redundant information for a single source

$$I_\cap(T; \boldsymbol{A}) = I(T; \boldsymbol{A}) = f\big((T, \boldsymbol{A})\big)$$

- Extending this dependence to multiple marginal vectors

$$I_\cap(T; \boldsymbol{A}_1, \ldots, \boldsymbol{A}_k) = f\big((T, \boldsymbol{A}_1), \ldots, (T, \boldsymbol{A}_k)\big)$$

- Many collections of marginal vectors are equivalent due to Axiom 3, e.g.

$$I_\cap\big((T, S_1), (T, S_1, S_2)\big) = I_\cap\big((T, S_1)\big)$$

Domain of the redundancy function for vectors

- Axiom 2 specifies the function form of redundant information for a single source

$$I_\cap(T; \boldsymbol{A}) = I(T; \boldsymbol{A}) = f\big((T, \boldsymbol{A})\big)$$

- Extending this dependence to multiple marginal vectors

$$I_\cap(T; \boldsymbol{A}_1, \dots, \boldsymbol{A}_k) = f\big((T, \boldsymbol{A}_1), \dots, (T, \boldsymbol{A}_k)\big)$$

- Many collections of marginal vectors are equivalent due to Axiom 3, e.g.

$$I_\cap\big((T, S_1), (T, S_1, S_2)\big) = I_\cap\big((T, S_1)\big)$$

- Unique collections are the antichains of the lattice of marginal vectors

Domain of the redundancy function for vectors

- Axiom 2 specifies the function form of redundant information for a single source

$$I_\cap(T;\boldsymbol{A}) = I(T;\boldsymbol{A}) = f\big((T,\boldsymbol{A})\big)$$

- Extending this dependence to multiple marginal vectors

$$I_\cap(T;\boldsymbol{A}_1,\ldots,\boldsymbol{A}_k) = f\big((T,\boldsymbol{A}_1),\ldots,(T,\boldsymbol{A}_k)\big)$$

- Many collections of marginal vectors are equivalent due to Axiom 3, e.g.

$$I_\cap\big((T,S_1),(T,S_1,S_2)\big) = I_\cap\big((T,S_1)\big)$$

- Unique collections are the antichains of the lattice of marginal vectors
- Antichains: $\{(T,\emptyset)\}$, $\{(T,S_1)\}$, $\{(T,S_2)\}$, $\{(T,S_1,S_2)\}$ and $\{(T,S_1),(T,S_2)\}$

Redundancy lattice for marginal vectors

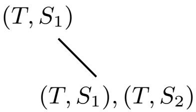
- Axiom 3 also defines a partial order $\preccurlyeq_\cap$ over these antichains

Redundancy lattice for marginal vectors

- Axiom 3 also defines a partial order $\preccurlyeq_\cap$ over these antichains
  - $I_\cap\big((T,S_1),(T,S_2)\big) \leq I_\cap\big((T,S_1)\big)$

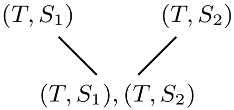
Redundancy lattice for marginal vectors

- Axiom 3 also defines a partial order $\preccurlyeq_\cap$ over these antichains
  - $I_\cap\big((T,S_1),(T,S_2)\big) \le I_\cap\big((T,S_1)\big)$



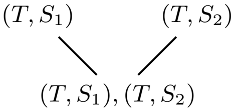
Redundancy lattice for marginal vectors

- Axiom 3 also defines a partial order $\preccurlyeq_\cap$ over these antichains
  - $I_\cap\big((T,S_1),(T,S_2)\big) \leq I_\cap\big((T,S_1)\big)$



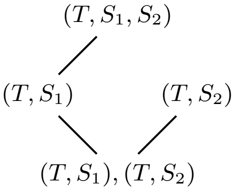
Redundancy lattice for marginal vectors

- Axiom 3 also defines a partial order $\preccurlyeq_\cap$ over these antichains
  - $I_\cap\big((T,S_1),(T,S_2)\big) \leq I_\cap\big((T,S_1)\big)$
  - $I_\cap\big((T,S_1)\big) = I_\cap\big((T,S_1),(T,S_1,S_2)\big) \leq I_\cap\big((T,S_1,S_2)\big)$



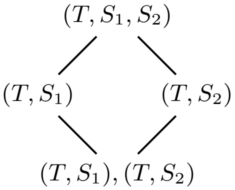
Redundancy lattice for marginal vectors

- Axiom 3 also defines a partial order $\preccurlyeq_\cap$ over these antichains
  - $I_\cap\big((T,S_1),(T,S_2)\big) \leq I_\cap\big((T,S_1)\big)$
  - $I_\cap\big((T,S_1)\big) = I_\cap\big((T,S_1),(T,S_1,S_2)\big) \leq I_\cap\big((T,S_1,S_2)\big)$



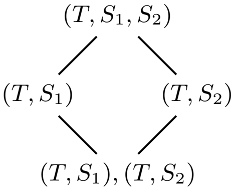
Redundancy lattice for marginal vectors

- Axiom 3 also defines a partial order $\preccurlyeq_\cap$ over these antichains
  - $I_\cap\big((T,S_1),(T,S_2)\big) \leq I_\cap\big((T,S_1)\big)$
  - $I_\cap\big((T,S_1)\big) = I_\cap\big((T,S_1),(T,S_1,S_2)\big) \leq I_\cap\big((T,S_1,S_2)\big)$



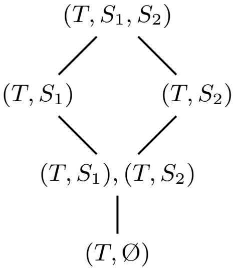
Redundancy lattice for marginal vectors

- Axiom 3 also defines a partial order $\preccurlyeq_\cap$ over these antichains
  - $I_\cap\big((T,S_1),(T,S_2)\big) \leq I_\cap\big((T,S_1)\big)$
  - $I_\cap\big((T,S_1)\big) = I_\cap\big((T,S_1),(T,S_1,S_2)\big) \leq I_\cap\big((T,S_1,S_2)\big)$
  - $I_\cap\big((T,\emptyset)\big) \leq I_\cap\big((T,S_1),(T,S_2)\big)$



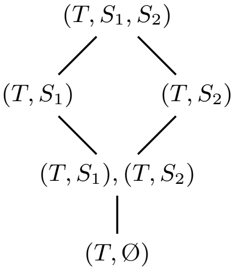
Redundancy lattice for marginal vectors

- Axiom 3 also defines a partial order $\preccurlyeq_\cap$ over these antichains
  - $I_\cap\big((T,S_1),(T,S_2)\big) \leq I_\cap\big((T,S_1)\big)$
  - $I_\cap\big((T,S_1)\big) = I_\cap\big((T,S_1),(T,S_1,S_2)\big) \leq I_\cap\big((T,S_1,S_2)\big)$
  - $I_\cap\big((T,\emptyset)\big) \leq I_\cap\big((T,S_1),(T,S_2)\big)$



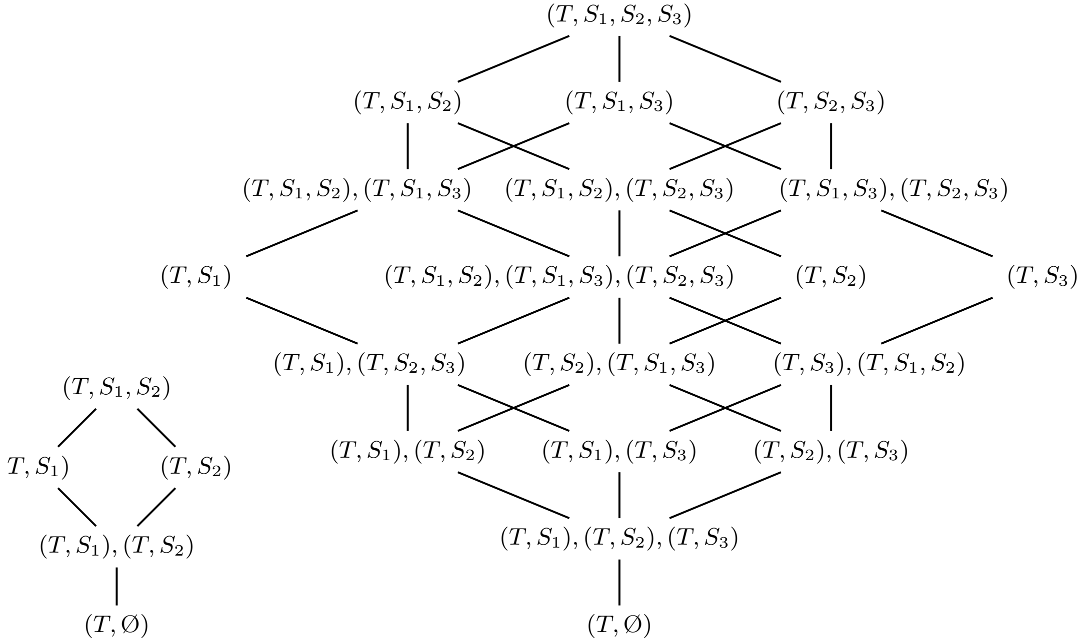
Redundancy lattice for marginal vectors

- Axiom 3 also defines a partial order $\preccurlyeq_\cap$ over these antichains
  - $I_\cap\big((T,S_1),(T,S_2)\big) \leq I_\cap\big((T,S_1)\big)$
  - $I_\cap\big((T,S_1)\big) = I_\cap\big((T,S_1),(T,S_1,S_2)\big) \leq I_\cap\big((T,S_1,S_2)\big)$
  - $I_\cap\big((T,\emptyset)\big) \leq I_\cap\big((T,S_1),(T,S_2)\big)$



- Redundant information only depends on the marginal distributions

$$I_\cap(T;\{S_1\},\{S_2\}) = f\big(P(T,S_1),P(T,S_2)\big)$$



(T, S1, S2, S3)
(T, S1, S2)
(T, S1, S3)
(T, S2, S3)
(T, S1, S2), (T, S1, S3)
(T, S1, S2), (T, S2, S3)
(T, S1, S3), (T, S2, S3)
(T, S1)
(T, S1, S2), (T, S1, S3), (T, S2, S3)
(T, S2)
(T, S3)
(T, S1), (T, S2, S3)
(T, S2), (T, S1, S3)
(T, S3), (T, S1, S2)
(T, S1), (T, S2)
(T, S1), (T, S3)
(T, S2), (T, S3)
(T, S1), (T, S2), (T, S3)
(T, ∅)
(T, S1, S2)
T, S1)
(T, S2)
(T, S1), (T, S2)
(T, ∅)

Functional dependence of redundant information

- If the redundant information depends on the full distribution, then

$$I_\cap(T;\{S_1\},\{S_2\}) = f\big(P(T,S_1,S_2)\big)$$

Functional dependence of redundant information

- If the redundant information depends on the full distribution, then

$$I_\cap(T; \{S_1\}, \{S_2\}) = f\big(P(T, S_1, S_2)\big)$$

  – But the joint source clearly also depends on the full distribution

$$I_\cap(T; \{S_1, S_2\}) = f\big(P(T, S_1, S_2)\big)$$

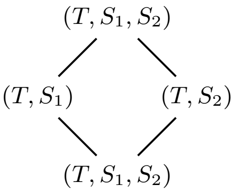
Functional dependence of redundant information

- If the redundant information depends on the full distribution, then

$$I_\cap(T; \{S_1\}, \{S_2\}) = f\big(P(T, S_1, S_2)\big)$$

  - But the joint source clearly also depends on the full distribution

$$I_\cap(T; \{S_1, S_2\}) = f\big(P(T, S_1, S_2)\big)$$



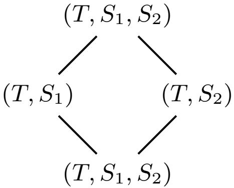
Functional dependence of redundant information

- If the redundant information depends on the full distribution, then

$$I_\cap(T;\{S_1\},\{S_2\}) = f\big(P(T,S_1,S_2)\big)$$

  – But the joint source clearly also depends on the full distribution

$$I_\cap(T;\{S_1,S_2\}) = f\big(P(T,S_1,S_2)\big)$$



- Redundant information only depends on the marginal distribution

$$I_\cap(T;\boldsymbol{A}_1,\ldots,\boldsymbol{A}_k) = f\big(P(T,\boldsymbol{A}_1),\ldots,(T,\boldsymbol{A}_k)\big)$$

Section 2: The total information lattice

Total information

- $I_\cap$ quantifies the redundant information a collection of sources provide about T
- Can we instead quantify the total information $I_\cup$ the sources provide about T?

Total information

- $I_\cap$ quantifies the redundant information a collection of sources provide about T
- Can we instead quantify the total information $I_\cup$ the sources provide about T?
  - this only makes sense if the total information depends only on the marginals

$$I_\cup(T; \boldsymbol{A}_1, \ldots, \boldsymbol{A}_k) = f\big((T, \boldsymbol{A}_1), \ldots, (T, \boldsymbol{A}_k)\big)$$

Total information

- $I_\cap$ quantifies the redundant information a collection of sources provide about T
- Can we instead quantify the total information $I_\cup$ the sources provide about T?
  - this only makes sense if the total information depends only on the marginals

$$I_\cup(T; \boldsymbol{A}_1, \ldots, \boldsymbol{A}_k) = f\big((T, \boldsymbol{A}_1), \ldots, (T, \boldsymbol{A}_k)\big)$$

  - if we knew the full distribution, then just use the (joint) mutual information

Axioms for total information

Axioms

1. *Symmetry*: $I_\cup(T; \boldsymbol{A}_1, \ldots, \boldsymbol{A}_k)$ is invariant under permutations of the $\boldsymbol{A}_i$'s
2. *Self-redundancy*: $I_\cup(T; \boldsymbol{A}_i) = I(T; \boldsymbol{A})$
3. *Monotonicity*:
$$I_\cup(T; \boldsymbol{A}_1, \ldots, \boldsymbol{A}_k) \geq I_\cup(T; \boldsymbol{A}_1, \ldots, \boldsymbol{A}_{k-1})$$
with equality if $\boldsymbol{A}_{k-1} \supseteq \boldsymbol{A}_k$

Axioms for total information

Axioms

1. *Symmetry*: $I_\cup(T; \boldsymbol{A}_1, \ldots, \boldsymbol{A}_k)$ is invariant under permutations of the $\boldsymbol{A}_i$'s
2. *Self-redundancy*: $I_\cup(T; \boldsymbol{A}_i) = I(T; \boldsymbol{A})$
3. *Monotonicity*:
$$I_\cup(T; \boldsymbol{A}_1, \ldots, \boldsymbol{A}_k) \geq I_\cup(T; \boldsymbol{A}_1, \ldots, \boldsymbol{A}_{k-1})$$
with equality if $\boldsymbol{A}_{k-1} \supseteq \boldsymbol{A}_k$

- What are the different ways the sources can provide total information?

Axioms for total information

Axioms

1. *Symmetry*: $I_\cup(T; \boldsymbol{A}_1, \ldots, \boldsymbol{A}_k)$ is invariant under permutations of the $\boldsymbol{A}_i$'s
2. *Self-redundancy*: $I_\cup(T; \boldsymbol{A}_i) = I(T; \boldsymbol{A})$
3. *Monotonicity*:

$$I_\cup(T; \boldsymbol{A}_1, \ldots, \boldsymbol{A}_k) \geq I_\cup(T; \boldsymbol{A}_1, \ldots, \boldsymbol{A}_{k-1})$$

   with equality if $\boldsymbol{A}_{k-1} \supseteq \boldsymbol{A}_k$

- What are the different ways the sources can provide total information?
  - Answering this question corresponds to determining the domain of $I_\cup$

Domain of the total information function

- $I_\cup$ can be applied to any collection of sources

Domain of the total information function

- $I_\cup$ can be applied to any collection of sources
- Many collection are equivalent due to Axiom 3, e.g. since $\{S_1\} \subseteq \{S_1, S_2\}$,

$$I_\cup(T; \{S_1\}, \{S_1, S_2\}) = I_\cup(T; \{S_1, S_2\})$$

Domain of the total information function

- $I_\cup$ can be applied to any collection of sources
- Many collection are equivalent due to Axiom 3, e.g. since $\{S_1\} \subseteq \{S_1, S_2\}$,

$$I_\cup(T; \{S_1\}, \{S_1, S_2\}) = I_\cup(T; \{S_1, S_2\})$$

- The unique collection are those such that no source is a superset of any other

Domain of the total information function

- $I_\cup$ can be applied to any collection of sources
- Many collection are equivalent due to Axiom 3, e.g. since $\{S_1\} \subseteq \{S_1, S_2\}$,

$$I_\cup(T; \{S_1\}, \{S_1, S_2\}) = I_\cup(T; \{S_1, S_2\})$$

- The unique collection are those such that no source is a superset of any other
  - the unique collections are also the antichains of the inclusion lattice

Domain of the total information function

- $I_\cup$ can be applied to any collection of sources
- Many collection are equivalent due to Axiom 3, e.g. since $\{S_1\} \subseteq \{S_1, S_2\}$,

$$I_\cup(T; \{S_1\}, \{S_1, S_2\}) = I_\cup(T; \{S_1, S_2\})$$

- The unique collection are those such that no source is a superset of any other
  - the unique collections are also the antichains of the inclusion lattice
  - same domain as the redundancy functions $I_\cap$

Total information lattice

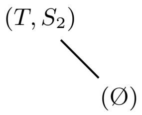
- The elements are the same, but the total information order $\preccurlyeq_{\cup}$ is different

Total information lattice

- The elements are the same, but the total information order $\preccurlyeq_{\cup}$ is different
  - $I_{\cup}(T;\{S_1\},\{S_2\}) \geq I_{\cup}(T;\{S_1\}) \geq I_{\cup}(T;\emptyset)$ follows directly from Axiom 3

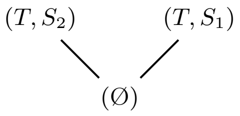
Total information lattice

- The elements are the same, but the total information order $\preccurlyeq_\cup$ is different
  - $I_\cup(T;\{S_1\},\{S_2\}) \geq I_\cup(T;\{S_1\}) \geq I_\cup(T;\emptyset)$ follows directly from Axiom 3



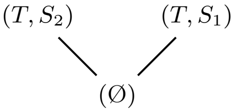
Total information lattice

- The elements are the same, but the total information order $\preccurlyeq_\cup$ is different
  - $I_\cup(T;\{S_1\},\{S_2\}) \geq I_\cup(T;\{S_1\}) \geq I_\cup(T;\emptyset)$ follows directly from Axiom 3



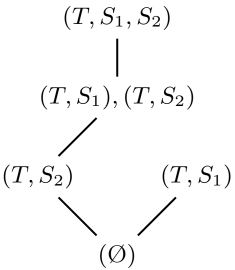
Total information lattice

- The elements are the same, but the total information order $\preccurlyeq_{\cup}$ is different
  - $I_{\cup}(T;\{S_1\},\{S_2\}) \geq I_{\cup}(T;\{S_1\}) \geq I_{\cup}(T;\emptyset)$ follows directly from Axiom 3
  - $I_{\cup}(T;\{S_1,S_2\}) = I_{\cup}(T;\{S_1\},\{S_2\},\{S_1,S_2\}) \geq I_{\cup}(T;\{S_1\},\{S_2\})$ since $\{S_1\} \subseteq \{S_1,S_2\}$ and $\{S_2\} \subseteq \{S_1,S_2\}$



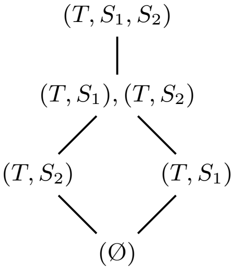
Total information lattice

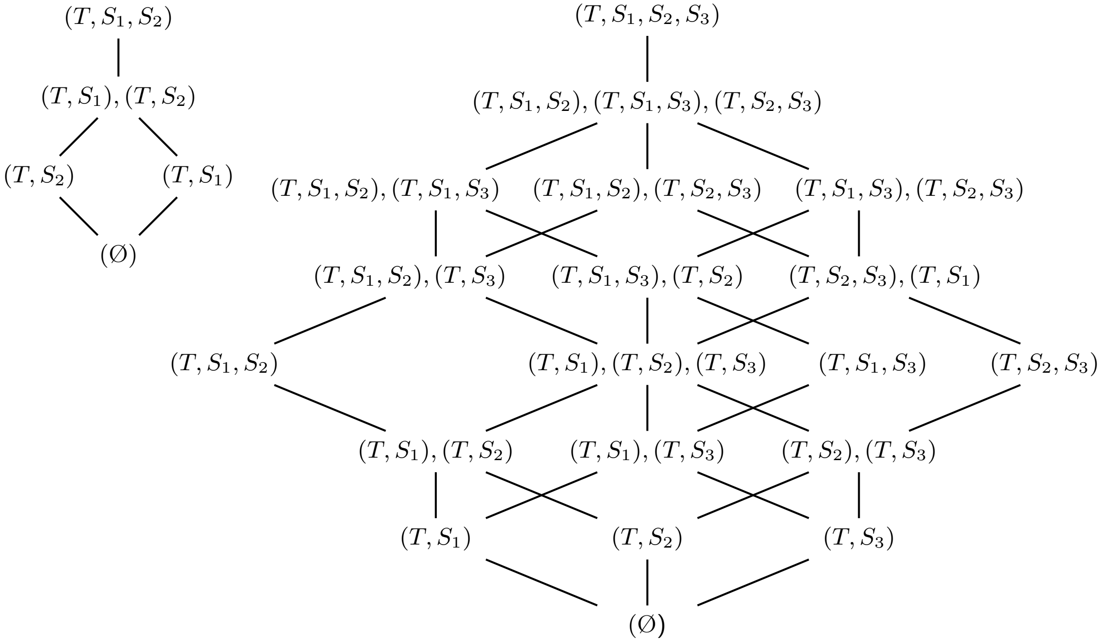
- The elements are the same, but the total information order $\preccurlyeq_\cup$ is different
  - $I_\cup(T;\{S_1\},\{S_2\}) \geq I_\cup(T;\{S_1\}) \geq I_\cup(T;\emptyset)$ follows directly from Axiom 3
  - $I_\cup(T;\{S_1,S_2\}) = I_\cup(T;\{S_1\},\{S_2\},\{S_1,S_2\}) \geq I_\cup(T;\{S_1\},\{S_2\})$ since $\{S_1\} \subseteq \{S_1,S_2\}$ and $\{S_2\} \subseteq \{S_1,S_2\}$



Total information lattice

- The elements are the same, but the total information order $\preccurlyeq_\cup$ is different
  - $I_\cup(T;\{S_1\},\{S_2\}) \geq I_\cup(T;\{S_1\}) \geq I_\cup(T;\emptyset)$ follows directly from Axiom 3
  - $I_\cup(T;\{S_1,S_2\}) = I_\cup(T;\{S_1\},\{S_2\},\{S_1,S_2\}) \geq I_\cup(T;\{S_1\},\{S_2\})$ since $\{S_1\} \subseteq \{S_1,S_2\}$ and $\{S_2\} \subseteq \{S_1,S_2\}$





(T, S1, S2)
(T, S1), (T, S2)
(T, S2)
(T, S1)
(∅)
(T, S1, S2, S3)
(T, S1, S2), (T, S1, S3), (T, S2, S3)
(T, S1, S2), (T, S1, S3)
(T, S1, S2), (T, S2, S3)
(T, S1, S3), (T, S2, S3)
(T, S1, S2), (T, S3)
(T, S1, S3), (T, S2)
(T, S2, S3), (T, S1)
(T, S1, S2)
(T, S1), (T, S2), (T, S3)
(T, S1, S3)
(T, S2, S3)
(T, S1), (T, S2)
(T, S1), (T, S3)
(T, S2), (T, S3)
(T, S1)
(T, S2)
(T, S3)
(∅)

Higher-order networks

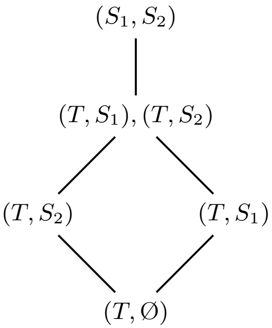
- Total information order $\preccurlyeq_{\cup}$ intuition aligns with higher-order networks

Higher-order networks

- Total information order $\preccurlyeq_\cup$ intuition aligns with higher-order networks
- Same order as the ordering of simplicial complexes containing T

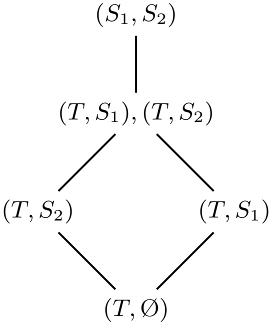
Higher-order networks

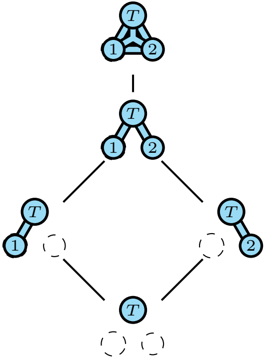
- Total information order $\preccurlyeq_\cup$ intuition aligns with higher-order networks
- Same order as the ordering of simplicial complexes containing T



Higher-order networks

- Total information order $\preccurlyeq_{\cup}$ intuition aligns with higher-order networks
- Same order as the ordering of simplicial complexes containing T





Connection between redundant and total information

- Redundancy $I_\cap$ and union information $I_\cup$ are dual concepts
- Union information order $\preccurlyeq_\cup$ intuition aligns with higher-order networks

Connection between redundant and total information

- Redundancy $I_\cap$ and union information $I_\cup$ are dual concepts
- Union information order $\preccurlyeq_\cup$ intuition aligns with higher-order networks
- Are the orders $\preccurlyeq_\cap$ and $\preccurlyeq_\cup$ connected by inclusion-exclusion?

$$I_\cup(T;S_1,S_2) = I(T;S_1) + I(T;S_2) - I_\cap(T;S_1,S_2)$$

Connection between redundant and total information

- Redundancy $I_\cap$ and union information $I_\cup$ are dual concepts
- Union information order $\preccurlyeq_\cup$ intuition aligns with higher-order networks
- Are the orders $\preccurlyeq_\cap$ and $\preccurlyeq_\cup$ connected by inclusion-exclusion?

$$I_\cup(T;S_1,S_2) = I(T;S_1) + I(T;S_2) - I_\cap(T;S_1,S_2)$$

  - min and max based approaches are connected

Connection between redundant and total information

- Redundancy $I_\cap$ and union information $I_\cup$ are dual concepts
- Union information order $\preccurlyeq_\cup$ intuition aligns with higher-order networks
- Are the orders $\preccurlyeq_\cap$ and $\preccurlyeq_\cup$ connected by inclusion-exclusion?

$$I_\cup(T;S_1,S_2) = I(T;S_1) + I(T;S_2) - I_\cap(T;S_1,S_2)$$

  - min and max based approaches are connected
  - the Blackwell orders are not connected

Connection between redundant and total information

- Redundancy $I_\cap$ and union information $I_\cup$ are dual concepts
- Union information order $\preccurlyeq_\cup$ intuition aligns with higher-order networks
- Are the orders $\preccurlyeq_\cap$ and $\preccurlyeq_\cup$ connected by inclusion-exclusion?

$$I_\cup(T; S_1, S_2) = I(T; S_1) + I(T; S_2) - I_\cap(T; S_1, S_2)$$

  - min and max based approaches are connected
  - the Blackwell orders are not connected

- Monotone $I_\cap$ requires that

$$I(T; S_1, S_2) \geq I(T; S_1) + I(T; S_2) - I_\cap(T; S_1, S_2)$$

Connection between redundant and total information

- Redundancy $I_\cap$ and union information $I_\cup$ are dual concepts
- Union information order $\preccurlyeq_\cup$ intuition aligns with higher-order networks
- Are the orders $\preccurlyeq_\cap$ and $\preccurlyeq_\cup$ connected by inclusion-exclusion?

$$I_\cup(T; S_1, S_2) = I(T; S_1) + I(T; S_2) - I_\cap(T; S_1, S_2)$$

  - min and max based approaches are connected
  - the Blackwell orders are not connected

- Monotone $I_\cap$ requires that

$$I(T; S_1, S_2) \geq I(T; S_1) + I(T; S_2) - I_\cap(T; S_1, S_2)$$

  - no assumption about inclusion-exclusion here

Total monotonicity

- The partial information atoms I_∂ are non-negative iff $I_\cap$ is totally monotone

$$I_\cap(T; \bigvee_{1\leq j\leq k} \alpha_j) \geq \sum_{\emptyset\neq J\subseteq\{1,\ldots,k\}} (-1)^{|J|-1} I_\cap(T; \bigwedge_{j\in J} \alpha_j)$$

Total monotonicity

- The partial information atoms I_∂ are non-negative iff $I_\cap$ is totally monotone

$$I_\cap(T; \bigvee_{1\leq j\leq k} \alpha_j) \geq \sum_{\emptyset\neq J\subseteq\{1,\ldots,k\}} (-1)^{|J|-1} I_\cap(T; \bigwedge_{j\in J} \alpha_j)$$

  - for $\alpha_1 = \{S_1\}$, $\alpha_2 = \{S_2\}$ and $\alpha_2 = \{S_2\}$

$$\begin{aligned} I\big(T; (S_1, S_2, S_3)\big) \geq\ & I(T; S_1) + I(T; S_2) + I(T; S_3) \\ & - I_\cap(T; S_1, S_2) - I_\cap(T; S_1, S_3) - I_\cap(T; S_2, S_3) \\ & + I_\cap(T; S_1, S_2, S_3) \end{aligned}$$

Total monotonicity

- The partial information atoms I_∂ are non-negative iff $I_\cap$ is totally monotone

$$I_\cap(T; \bigvee_{1 \le j \le k} \alpha_j) \ge \sum_{\emptyset \ne J \subseteq \{1,\dots,k\}} (-1)^{|J|-1} I_\cap(T; \bigwedge_{j \in J} \alpha_j)$$

  - for $\alpha_1 = \{S_1\}$, $\alpha_2 = \{S_2\}$ and $\alpha_2 = \{S_2\}$

$$\begin{aligned} I\big(T;(S_1,S_2,S_3)\big) \ge\; & I(T;S_1) + I(T;S_2) + I(T;S_3) \\ & - I_\cap(T;S_1,S_2) - I_\cap(T;S_1,S_3) - I_\cap(T;S_2,S_3) \\ & + I_\cap(T;S_1,S_2,S_3) \end{aligned}$$

  - for $\alpha_1 = \{S_1, S_2\}$, $\alpha_2 = \{S_1, S_3\}$ and $\alpha_2 = \{S_2, S_3\}$

$$\begin{aligned} I\big(T;(S_1,S_2,S_3)\big) \ge\; & I\big(T;(S_1,S_2)\big) + I\big(T;(S_1,S_3)\big) + I\big((T;(S_2,S_3)\big) \\ & - I_\cap\big(T;(S_1,S_2),(S_1,S_3)\big) - I_\cap\big(T;(S_1,S_2),(S_2,S_3)\big) - I_\cap\big(T;(S_1,S_3),(S_2,S_3)\big) \\ & + I_\cap\big(T;(S_1,S_2),(S_1,S_3),(S_2,S_3)\big) \end{aligned}$$

Total monotonicity

- The partial information atoms I_∂ are non-negative iff $I_\cap$ is totally monotone

$$I_\cap(T; \bigvee_{1\le j\le k} \alpha_j) \ge \sum_{\emptyset\neq J\subseteq\{1,\ldots,k\}} (-1)^{|J|-1} I_\cap(T; \bigwedge_{j\in J} \alpha_j)$$

  - for $\alpha_1 = \{S_1\}$, $\alpha_2 = \{S_2\}$ and $\alpha_2 = \{S_2\}$

$$\begin{aligned} I\big(T;(S_1,S_2,S_3)\big) \ge\ & I(T;S_1) + I(T;S_2) + I(T;S_3) \\ & - I_\cap(T;S_1,S_2) - I_\cap(T;S_1,S_3) - I_\cap(T;S_2,S_3) \\ & + I_\cap(T;S_1,S_2,S_3) \end{aligned}$$

  - for $\alpha_1 = \{S_1, S_2\}$, $\alpha_2 = \{S_1, S_3\}$ and $\alpha_2 = \{S_2, S_3\}$

$$\begin{aligned} I\big(T;(S_1,S_2,S_3)\big) \ge\ & I\big(T;(S_1,S_2)\big) + I\big(T;(S_1,S_3)\big) + I\big((T;(S_2,S_3)\big) \\ & - I_\cap\big(T;(S_1,S_2),(S_1,S_3)\big) - I_\cap\big(T;(S_1,S_2),(S_2,S_3)\big) - I_\cap\big(T;(S_1,S_3),(S_2,S_3)\big) \\ & + I_\cap\big(T;(S_1,S_2),(S_1,S_3),(S_2,S_3)\big) \end{aligned}$$

- Total monotonicity for $I_\cap$ fails and when a connected $I_\cup$ fails to be monotonic

Conclusions

Redundant information $I_\cap$ only depends of the marginal distributions

Conclusions

Redundant information $I_\cap$ only depends of the marginal distributions

Redundant information $I_\cap$ is only one side of the information decomposition problem

Conclusions

Redundant information $I_\cap$ only depends of the marginal distributions

Redundant information $I_\cap$ is only one side of the information decomposition problem

- We also need to consider the union information $I_\cup$

Conclusions

Redundant information $I_\cap$ only depends of the marginal distributions

Redundant information $I_\cap$ is only one side of the information decomposition problem

– We also need to consider the union information $I_\cup$

– Easier to consider the monotonicity of $I_\cup$ than total monotonicity of $I_\cup$