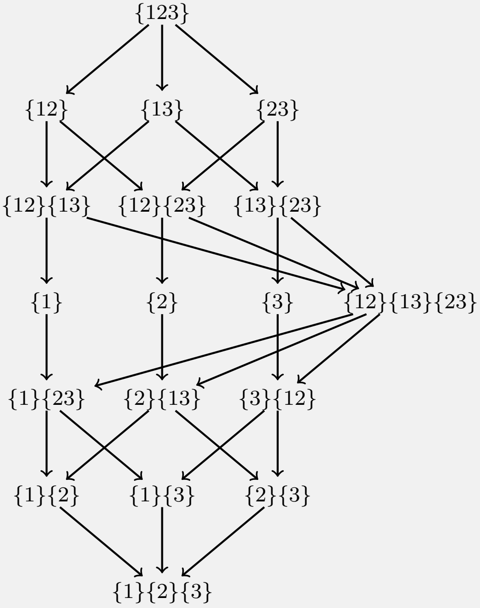
Pointwise Partial Information Decomposition Using Specificity and Ambiguity Lattices

MPI DS, Göttingen

Conor Finn
Joseph Lizier

December 13, 2017

THE UNIVERSITY OF SYDNEY



Unique, redundant and synergistic information

Consider three random variables S_1, S_2 and T

- Aim: predict T using S_1 and S_2
- Several types of information:

1. **Unique information** $U(S_1 \setminus S_2 \to T)$
2. **Redundant information** $R(S_1, S_2 \to T)$
3. **Complementary 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

Bivariate information decomposition

In general, all types of information are present

- Mutual information captures

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

- Joint mutual information captures

$$I(S_{1,2};T) = R(S_1,S_2 \to T) + U(S_1 \setminus S_2 \to T) + U(S_2 \setminus S_1 \to T) + C(S_1,S_2 \to T)$$

- Three equations with four unknowns

AND

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	1

Partial information decomposition

- Axiomatic framework extending this decomposition to arbitrary number of source

Axioms: PID

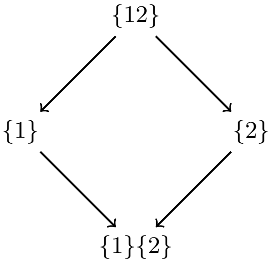
(1) *Symmetry*: $R(S_1, \ldots, S_n \rightarrow T)$ is invariant under permutations of the S_i's

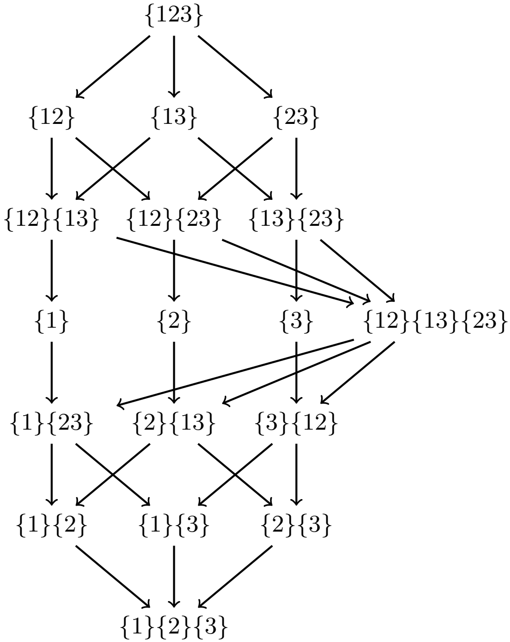
(2) *Monotonicity*: $R(S_1, \ldots, S_n \rightarrow T) \leq R(S_1, \ldots, S_{n-1} \rightarrow T)$

(3) *Self-redundancy*: $R(S_i \rightarrow T) = I(S_i; T)$

- Yields a **redundancy lattice**
- No well-accepted, compatible definition redundant information
- Only consistent for one target variable (no target chain rule)

Redundancy lattice





Pointwise information theory

- The mutual information
$$I(X;Y) = \sum_{x\in\mathcal{X},\, y\in\mathcal{Y}} \log \frac{(x,y)}{p(x)p(y)} \geq 0$$

- Woodward (1953) noted the average form of “tempts one to enquire into other simpler methods of derivation [of the per state information]”.

Pointwise information theory

- The mutual information

$$I(X;Y) = \sum_{x\in\mathcal{X},\, y\in\mathcal{Y}} \log \frac{(x,y)}{p(x)p(y)} \geq 0$$

- Woodward (1953) noted the average form of "tempts one to enquire into other simpler methods of derivation [of the per state information]".

- Using two postulate, Woodward derived the **pointwise** mutual information

$$i(x;y) = \log \frac{p(x,y)}{p(x)p(y)} \ngeq 0$$

- This was later done more formally by Fano (1961) using four postulates

- Corollaries: (average) mutual information, pointwise entropy and (Shannon) entropy

Bivariate pointwise information decomposition

- Pointwise decomposition for each realisation

$$
\begin{aligned}
i(s_1;t) &= r(s_1,s_2 \to t) + u(s_1 \setminus s_2 \to t) \\
i(s_2;t) &= r(s_1,s_2 \to t) + u(s_2 \setminus s_1 \to t) \\
i(s_{1,2};t) &= r(s_1,s_2 \to t) + u(s_1 \setminus s_2 \to t) + u(s_2 \setminus s_1 \to t) + c(s_1,s_2 \to t)
\end{aligned}
$$

Bivariate pointwise information decomposition

- Pointwise decomposition for each realisation

$$i(s_1;t) = r(s_1, s_2 \to t) + u(s_1 \setminus s_2 \to t)$$

$$i(s_2;t) = r(s_1, s_2 \to t) + u(s_2 \setminus s_1 \to t)$$

$$i(s_{1,2};t) = r(s_1, s_2 \to t) + u(s_1 \setminus s_2 \to t) + u(s_2 \setminus s_1 \to t) + c(s_1, s_2 \to t)$$

- Should be able to take the expectation over all realisations

$$R(S_1, S_2 \to T) = \big\langle r(s_1, s_2 \to t) \big\rangle \qquad U(S_1 \setminus S_2 \to T) = \big\langle u(s_1 \setminus s_2 \to t) \big\rangle$$

$$C(S_1, S_2 \to T) = \big\langle c(s_1, s_2 \to t) \big\rangle \qquad U(S_2 \setminus S_1 \to T) = \big\langle u(s_2 \setminus s_1 \to t) \big\rangle$$

- This should recover the (average) information decomposition

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

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

$$I(S_{1,2};T) = R(S_1, S_2 \to T) + U(S_1 \setminus S_2 \to T) + U(S_2 \setminus S_1 \to T) + C(S_1, S_2 \to T)$$

Motivation: PwUnq

- Consider PwUnq from Finn and Lizier (2017b)

p	s_1	s_2	t	
1/4	0	1	1	
1/4	1	0	1	
1/4	0	2	2	
1/4	2	0	2	
Expected values				

Motivation: PwUnq

- Consider PwUnq from Finn and Lizier (2017b)

p	s_1	s_2	t	i_1	i_2	$i_{1,2}$	
1/4	0	1	1	0	1	1	
1/4	1	0	1	1	0	1	
1/4	0	2	2	0	1	1	
1/4	2	0	2	1	0	1	
Expected values				1/2	1/2	1	

Motivation: PwUnq

- Consider PwUnq from Finn and Lizier (2017b)

p	s_1	s_2	t	i_1	i_2	$i_{1,2}$	r
1/4	0	1	1	0	1	1	0
1/4	1	0	1	1	0	1	0
1/4	0	2	2	0	1	1	0
1/4	2	0	2	1	0	1	0
Expected values				1/2	1/2	1	0

Motivation: PwUnq

- Consider PwUnq from Finn and Lizier (2017b)

p	s_1	s_2	t	i_1	i_2	$i_{1,2}$	r	u_1	u_2	c
1/4	0	1	1	0	1	1	0	0	1	0
1/4	1	0	1	1	0	1	0	1	0	0
1/4	0	2	2	0	1	1	0	0	1	0
1/4	2	0	2	1	0	1	0	1	0	0
Expected values				1/2	1/2	1	0	1/2	1/2	0

Motivation: PwUnq

- Consider PwUnq from Finn and Lizier (2017b)

p	s_1	s_2	t	i_1	i_2	$i_{1,2}$	r	u_1	u_2	c
1/4	0	1	1	0	1	1	0	0	1	0
1/4	1	0	1	1	0	1	0	1	0	0
1/4	0	2	2	0	1	1	0	0	1	0
1/4	2	0	2	1	0	1	0	1	0	0
Expected values				1/2	1/2	1	0	1/2	1/2	0

- According to I_{min} Williams and Beer (2010), $\widetilde{UI}$ of Bertschinger et al. (2014), S_{VK} of Griffith and Koch (2014) and I_{red} of Harder et al. (2013)

$$R = \langle r \rangle = 1/2 \text{ bit} \neq 0 \text{ bit}$$

- We refer to as the **pointwise unique problem**

Pointwise partial information decomposition

Axioms: PPID

(1) *Symmetry*: $r(s_1, \ldots, s_n \to t)$ is invariant under permutations of the s_i's

(2) *Monotonicity*: $r(s_1, \ldots, s_n \to t) \leq r(s_1, \ldots, s_{n-1} \to t)$

(3) *Self-redundancy*: $r(s_i \to t) = i(s_i; t)$

- Would yield a redundancy lattice for every joint realisation

Pointwise partial information decomposition

Axioms: PPID

(1) *Symmetry*: $r(s_1, \ldots, s_n \to t)$ is invariant under permutations of the s_i's

(2) *Monotonicity*: $r(s_1, \ldots, s_n \to t) \leq r(s_1, \ldots, s_{n-1} \to t)$

(3) *Self-redundancy*: $r(s_i \to t) = i(s_i; t)$

- Would yield a redundancy lattice for every joint realisation
- Problems:
  1. Pointwise mutual information is not non-negative
  2. Still no clear definition of redundant information
- PPID needs to overcome this lack of non-negativity

Probability mass exclusions

- By definition, the pointwise mutual information provided by s about t

Surprise of the prior $\quad p(t) \longrightarrow p(t|s) \quad$ Surprise of the posterior

Probability mass exclusions

- By definition, the pointwise mutual information provided by s about t

  Surprise of the prior $\quad p(t) \longrightarrow p(t|s) \quad$ Surprise of the posterior

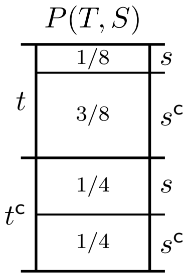
- **Finn and Lizier (2017a)**: this change is derived from probability mass exclusions

Probability mass exclusions

- By definition, the pointwise mutual information provided by s about t

Surprise of the prior $\quad p(t) \longrightarrow p(t|s) \quad$ Surprise of the posterior

- **Finn and Lizier (2017a)**: this change is derived from probability mass exclusions

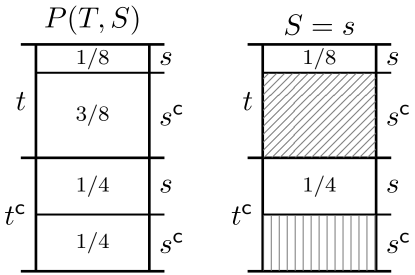


Probability mass exclusions

- By definition, the pointwise mutual information provided by s about t

Surprise of the prior $\quad p(t) \longrightarrow p(t|s) \quad$ Surprise of the posterior

- **Finn and Lizier (2017a)**: this change is derived from probability mass exclusions

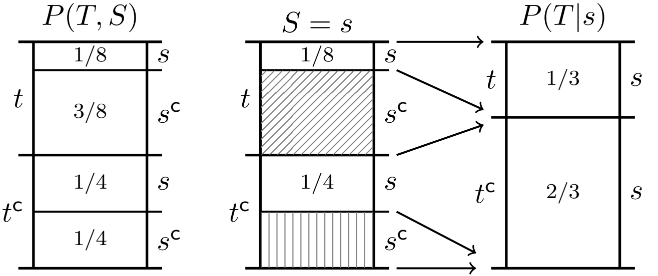


Probability mass exclusions

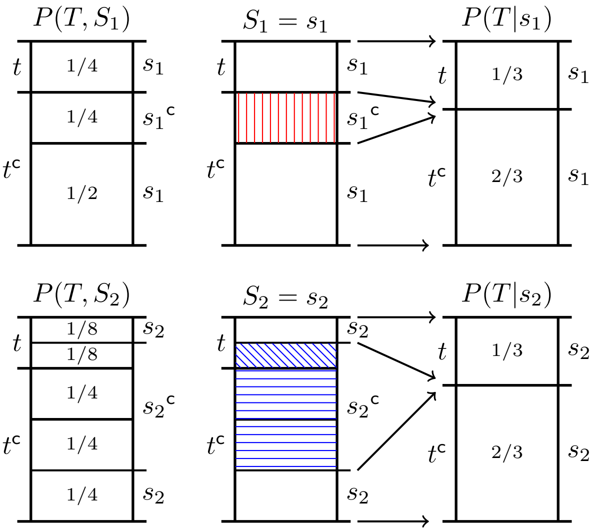
- By definition, the pointwise mutual information provided by s about t

Surprise of the prior $\quad p(t) \longrightarrow p(t|s) \quad$ Surprise of the posterior

- **Finn and Lizier (2017a)**: this change is derived from probability mass exclusions



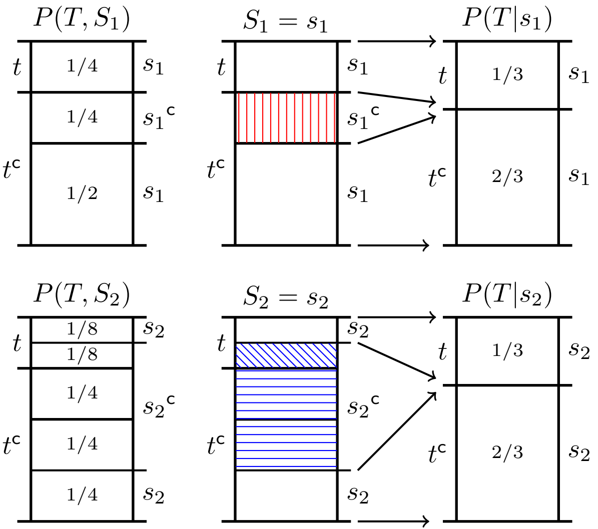
Motivation for this approach



- The exclusions differ, but yet

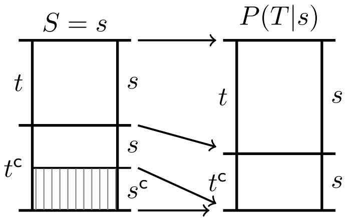
  $i(t; s_1) = i(t; s_2) = {}^4\!/_3$ bit

Motivation for this approach

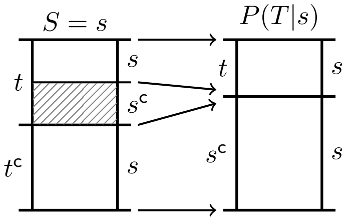


- The exclusions differ, but yet $i(t; s_1) = i(t; s_2) = {}^4\!/_3$ bit
- Pointwise MI is not injective
- Same info $\leftrightarrow$ same exclusions

Two types of exclusions

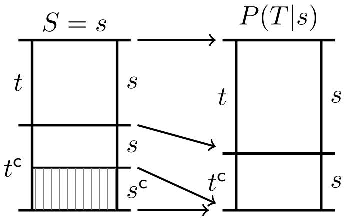


Purely informative exclusion

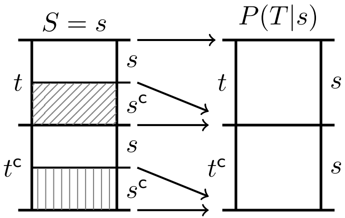


Purely misinformative exclusions

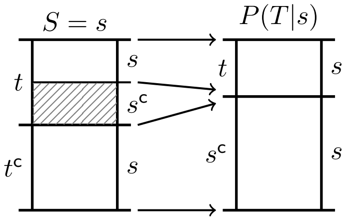
Two types of exclusions



Purely informative exclusion



General case



Purely misinformative exclusions

- Idea: split the pointwise MI into two components

$$i(s \to t) = i^{+}(s \to t) - i^{-}(s \to t)$$

Postulates for the decomposition

Postulate 1 The information provided by s about t can be decomposed into two non-negative components,

$$i(s;t) = i^{+}(s \to t) - i^{-}(s \to t).$$

Postulate 2 Each component satisfy a chain rule,

$$i_{+}(s_{1,2} \to t) = i_{+}(s_1 \to t) + i_{+}(s_2 \to t|s_1),$$
$$i_{-}(s_{1,2} \to t) = i_{-}(s_1 \to t) + i_{-}(s_2 \to t|s_1).$$

Postulate 3 The components $i^{+}(s \to t)$ and $i^{+}(s \to t)$ are continuous, monotonically increasing functions the informative and misinformative exclusions, respectively.

Postulates for the decomposition

Postulate 1 The information provided by s about t can be decomposed into two non-negative components,

$$i(s;t) = i^{+}(s \to t) - i^{-}(s \to t).$$

Postulate 2 Each component satisfy a chain rule,

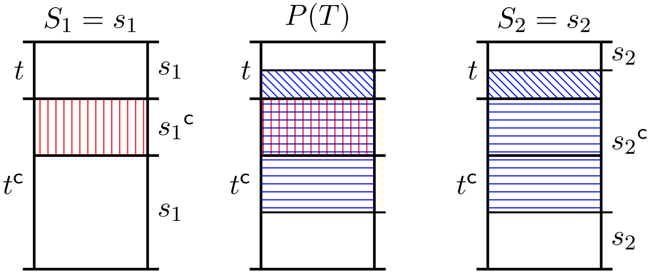
$$i_{+}(s_{1,2} \to t) = i_{+}(s_1 \to t) + i_{+}(s_2 \to t|s_1),$$
$$i_{-}(s_{1,2} \to t) = i_{-}(s_1 \to t) + i_{-}(s_2 \to t|s_1).$$

Postulate 3 The components $i^{+}(s \to t)$ and $i^{+}(s \to t)$ are continuous, monotonically increasing functions the informative and misinformative exclusions, respectively.

- Finn and Lizier (2017a) proved that

(**Specificity**) $i^{+}(s \to t) = h(s)$ $\qquad$ $i^{-}(s \to t) = h(s|t)$ (**Ambiguity**)

Specificity and ambiguity decomposition



$$i(s_1 \to t) = i(s_2 \to t) = \log {}^4\!/_3 \text{ bit}$$

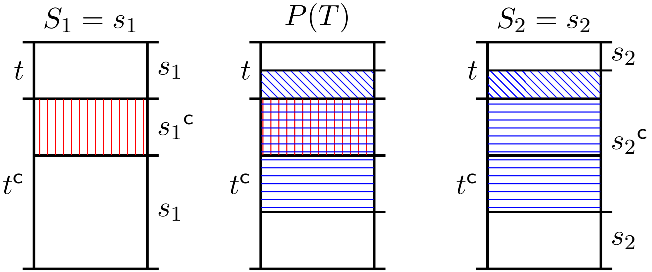
$i_+(s_1 \to t) = \log {}^4\!/_3$ bit

$i_+(s_2 \to t) = \log \frac{8}{3}$ bit

$i_-(s_1 \to t) = 0$ bit

$i_-(s_2 \to t) = 1$ bit

Specificity and ambiguity decomposition



$$i(s_1 \to t) = i(s_2 \to t) = \log {}^4\!/_3 \text{ bit}$$

$i_+(s_1 \to t) = \log {}^4\!/_3$ bit

$i_-(s_1 \to t) = 0$ bit

$i_+(s_2 \to t) = \log \frac{8}{3}$ bit

$i_-(s_2 \to t) = 1$ bit

$r^+(s_1, s_2 \to t) = \log {}^4\!/_3$ bit, $\quad u^+(s_2 \backslash s_2 \to t) = \log {}^4\!/_3$ bit $\quad u^-(s_2 \backslash s_2 \to t) = \log 1$ bit

PPID using specificity and ambiguity

- **Finn and Lizier (2017b)** proposes

Axioms: PPID using Specificity and Ambiguity

(1) *Symmetry*: $r^{\pm}(s_1, \ldots, s_n \to t)$ is invariant under permutations of the s_i's

(2) *Monotonicity*: $r^{\pm}(s_1, \ldots, s_n \to t) \leq r^{\pm}(s_1; \ldots; s_{n-1} \to t)$

(3) *Self-redundancy*: $r^{\pm}(s_i \to t) = i^{\pm}(s_i; t)$

- Yields a **specificity lattice** and an **ambiguity lattice** for every joint realisation
- Circumvents the non-negativity problem
- Require a measure of redundant specificity and redundant ambiguity

Pointwise redundant specificity and ambiguity

- The pointwise mutual information $i(s;t)$ does not depend on the apportionment of the probability mass exclusions within the complementary event t^{c}.

Pointwise redundant specificity and ambiguity

- The pointwise mutual information $i(s;t)$ does not depend on the apportionment of the probability mass exclusions within the complementary event t^{c}.

Axioms: PPID using Specificity and Ambiguity (cont.)

(4) *Two-event partition*: $r^{\pm}(t : s_1, \ldots, s_n)$ are functions of the probability measures on the two-event partitions $\mathcal{S}_1^{s_1} \times \mathcal{T}^t, \ldots, \mathcal{S}_n^{s_n} \times \mathcal{T}^t$

- Finn and Lizier (2017b) provides arguments justifying this in terms of Kelly gambling

Pointwise redundant specificity and ambiguity

- The pointwise mutual information $i(s;t)$ does not depend on the apportionment of the probability mass exclusions within the complementary event t^{c}.

Axioms: PPID using Specificity and Ambiguity (cont.)

(4) *Two-event partition*: $r^{\pm}(t : s_1, \ldots, s_n)$ are functions of the probability measures on the two-event partitions $\mathcal{S}_1^{s_1} \times \mathcal{T}^t, \ldots, \mathcal{S}_n^{s_n} \times \mathcal{T}^t$

- Finn and Lizier (2017b) provides arguments justifying this in terms of Kelly gambling
- Leads us to define the redundant specificity and redundant ambiguity

$$r^{+}_{\mathsf{min}}(s_1, \ldots, s_n \to t) = \min_{s_j} h(s_j) \qquad r^{-}_{\mathsf{min}}(s_1, \ldots, s_n \to t) = \min_{s_j} h(s_j|t)$$

- Upon recombination, this measure satisfies the target chain rule

$$r_{\mathsf{min}}\big(s_1, \ldots, s_n \to t_{1,2}\big) = r_{\mathsf{min}}\big(s_1, \ldots, s_n \to t_1\big) + r_{\mathsf{min}}\big(s_1, \ldots, s_n \to t_2|t_1\big),$$

Bivariate PPID using specificity and ambiguity

- Decomposition of both the specificity and ambiguity the for each realisation

$$
\begin{aligned}
i^{\pm}(s_1;t) &= r^{\pm}(s_1,s_2 \to t) + u^{\pm}(s_1 \setminus s_2 \to t) \\
i^{\pm}(s_2;t) &= r^{\pm}(s_1,s_2 \to t) + u^{\pm}(s_2 \setminus s_1 \to t) \\
i^{\pm}(s_{1,2};t) &= r^{\pm}(s_1,s_2 \to t) + u^{\pm}(s_1 \setminus s_2 \to t) + u^{\pm}(s_2 \setminus s_1 \to t) + c^{\pm}(s_1,s_2 \to t)
\end{aligned}
$$

Bivariate PPID using specificity and ambiguity

- Decomposition of both the specificity and ambiguity the for each realisation

$$
\begin{aligned}
i^{\pm}(s_1;t) &= r^{\pm}(s_1,s_2\to t) + u^{\pm}(s_1\setminus s_2\to t)\\
i^{\pm}(s_2;t) &= r^{\pm}(s_1,s_2\to t) + u^{\pm}(s_2\setminus s_1\to t)\\
i^{\pm}(s_{1,2};t) &= r^{\pm}(s_1,s_2\to t) + u^{\pm}(s_1\setminus s_2\to t) + u^{\pm}(s_2\setminus s_1\to t) + c^{\pm}(s_1,s_2\to t)
\end{aligned}
$$

- Taking the expectation yields

$$
\begin{aligned}
R^{\pm}(S_1,S_2\to T) &= \big\langle r^{\pm}(s_1,s_2\to t)\big\rangle & U^{\pm}(S_1\setminus S_2\to T) &= \big\langle u^{\pm}(s_1\setminus s_2\to t)\big\rangle\\
C^{\pm}(S_1,S_2\to T) &= \big\langle c^{\pm}(s_1,s_2\to t)\big\rangle & U^{\pm}(S_2\setminus S_1\to T) &= \big\langle u^{\pm}(s_2\setminus s_1\to t)\big\rangle
\end{aligned}
$$

- And we have a decomposition of average specificity and ambiguity

$$
\begin{aligned}
I^{\pm}(S_1;T) &= R^{\pm}(S_1,S_2\to T) + U^{\pm}(S_1\setminus S_2\to T)\\
I^{\pm}(S_2;T) &= R^{\pm}(S_1,S_2\to T) + U^{\pm}(S_2\setminus S_1\to T)\\
I^{\pm}(S_{1,2};T) &= R^{\pm}(S_1,S_2\to T) + U^{\pm}(S_1\setminus S_2\to T) + U^{\pm}(S_2\setminus S_1\to T) + C^{\pm}(S_1,S_2\to T)
\end{aligned}
$$

Bivariate PPID using specificity and ambiguity (cont.)

- Or recombine the specificity and ambiguity for pointwise information

$$
\begin{aligned}
r(s_1, s_2 \to t) &= r^+(s_1, s_2 \to t) \ - r^-(s_1, s_2 \to t) \\
u(s_1 \setminus s_2 \to t) &= u^+(s_1 \setminus s_2 \to t) - u^-(s_1 \setminus s_2 \to t) \\
u(s_2 \setminus s_1 \to t) &= u^+(s_2 \setminus s_1 \to t) - u^-(s_2 \setminus s_1 \to t) \\
c(s_1, s_2 \to t) &= c^+(s_1, s_2 \to t) \ - c^-(s_1, s_2 \to t)
\end{aligned}
$$

Bivariate PPID using specificity and ambiguity (cont.)

- Or recombine the specificity and ambiguity for pointwise information

$$r(s_1, s_2 \to t) = r^+(s_1, s_2 \to t) \; - r^-(s_1, s_2 \to t)$$
$$u(s_1 \setminus s_2 \to t) = u^+(s_1 \setminus s_2 \to t) - u^-(s_1 \setminus s_2 \to t)$$
$$u(s_2 \setminus s_1 \to t) = u^+(s_2 \setminus s_1 \to t) - u^-(s_2 \setminus s_1 \to t)$$
$$c(s_1, s_2 \to t) = c^+(s_1, s_2 \to t) \; - c^-(s_1, s_2 \to t)$$

- Which satisfy the PPID

$$i(s_1; t) = r(s_1, s_2 \to t) + u(s_1 \setminus s_2 \to t)$$
$$i(s_2; t) = r(s_1, s_2 \to t) + u(s_2 \setminus s_1 \to t)$$
$$i(s_{1,2}; t) = r(s_1, s_2 \to t) + u(s_1 \setminus s_2 \to t) + u(s_2 \setminus s_1 \to t) + c(s_1, s_2 \to t)$$

Bivariate PPID using specificity and ambiguity (cont.)

- Or recombine the specificity and ambiguity for pointwise information

$$r(s_1, s_2 \to t) = r^+(s_1, s_2 \to t) \; - r^-(s_1, s_2 \to t)$$
$$u(s_1 \backslash s_2 \to t) = u^+(s_1 \backslash s_2 \to t) - u^-(s_1 \backslash s_2 \to t)$$
$$u(s_2 \backslash s_1 \to t) = u^+(s_2 \backslash s_1 \to t) - u^-(s_2 \backslash s_1 \to t)$$
$$c(s_1, s_2 \to t) = c^+(s_1, s_2 \to t) \; - c^-(s_1, s_2 \to t)$$

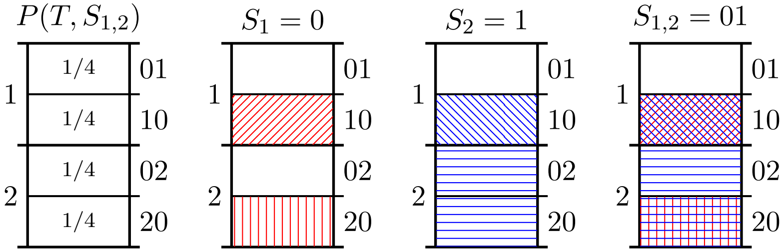
- Which satisfy the PPID

$$i(s_1; t) = r(s_1, s_2 \to t) + u(s_1 \backslash s_2 \to t)$$
$$i(s_2; t) = r(s_1, s_2 \to t) + u(s_2 \backslash s_1 \to t)$$
$$i(s_{1,2}; t) = r(s_1, s_2 \to t) + u(s_1 \backslash s_2 \to t) + u(s_2 \backslash s_1 \to t) + c(s_1, s_2 \to t)$$

- While taking the expectation yields the PID

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

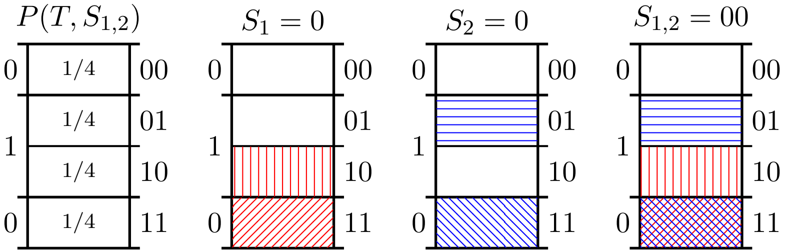
Example: PwUnq



p	s_1	s_2	t	i_1^+	i_1^-	i_2^+	i_2^-	$i_{1,2}^+$	$i_{1,2}^-$	r^+	u_1^+	u_2^+	c^+	r^-	u_1^-	u_2^-	c^-
1/4	0	1	1	1	1	2	1	2	1	1	0	1	0	1	0	0	0
1/4	1	0	1	2	1	1	1	2	1	1	1	0	0	1	0	0	0
1/4	0	2	2	1	1	2	1	2	1	1	0	1	0	1	0	0	0
1/4	2	0	2	2	1	1	1	2	1	1	1	0	0	1	0	0	0
Expected values				3/2	1	3/2	1	2	1	1	1/2	1/2	0	1	0	0	0

$R(S_1, S_2 \to T) = 0 \quad U(S_1 \setminus S_2 \to T) = 1/2 \quad U(S_2 \setminus S_1 \to T) = 1/2 \quad C(S_1, S_2 \to T) = 0$

Example: XOR



p	s_1	s_2	t	i_1^+	i_1^-	i_2^+	i_2^-	$i_{1,2}^+$	$i_{1,2}^-$	r^+	u_1^+	u_2^+	c^+	r^-	u_1^-	u_2^-	c^-
1/4	0	0	0	1	1	1	1	2	1	1	0	0	1	1	0	0	0
1/4	0	1	1	1	1	1	1	2	1	1	0	0	1	1	0	0	0
1/4	1	0	1	1	1	1	1	2	1	1	0	0	1	1	0	0	0
1/4	1	1	0	1	1	1	1	2	1	1	0	0	1	1	0	0	0
Expected values				1	1	1	1	2	1	1	0	0	1	1	0	0	0

$R(S_1, S_2 \to T) = 0$ $\quad U(S_1 \setminus S_2 \to T) = 0$ $\quad U(S_2 \setminus S_1 \to T) = 0$ $\quad C(S_1, S_2 \to T) = 1$

Comparison to Other Decompositions and Measures

- Similar to Ince (2017) but the monotonicity issue is dealt with in a principled way
- Similar to I_{min} of Williams and Beer (2010) but now fully pointwise
- Axiom 4 is similar to Assumption $(**)$ of Bertschinger et al. (2014), i.e. measure $\widetilde{UI}$
- This also makes it similar to S_{VK} of Griffith and Koch (2014)
- There is a target chain rule but no target monotonicity
- There is local positivity on the specificity and ambiguity lattices
- However, upon recombination and taking the expectation, the PI atoms can be negative

References

Nils Bertschinger, Johannes Rauh, Eckehard Olbrich, Jürgen Jost, and Nihat Ay. Quantifying unique information. *Entropy*, 16(4):2161–2183, 2014.

Robert Fano. *Transmission of Information*. The MIT Press, 1961.

Conor Finn and Joseph T. Lizier. Probability mass exclusions and the directed components of pointwise mutual information. 2017a. To appear.

Conor Finn and Joseph T. Lizier. Pointwise partial information decomposition using the specificity and ambiguity lattices. 2017b. To appear.

Virgil Griffith and Christof Koch. Quantifying synergistic mutual information. In Mikhail Prokopenko, editor, *Guided Self-Organization: Inception*, volume 9 of *Emergence, Complexity and Computation*, pages 159–190. Springer Berlin Heidelberg, 2014. ISBN 978-3-642-53733-2.

Malte Harder, Christoph Salge, and Daniel Polani. Bivariate measure of redundant information. *Physical Review E*, 87 (1):012130, 2013.

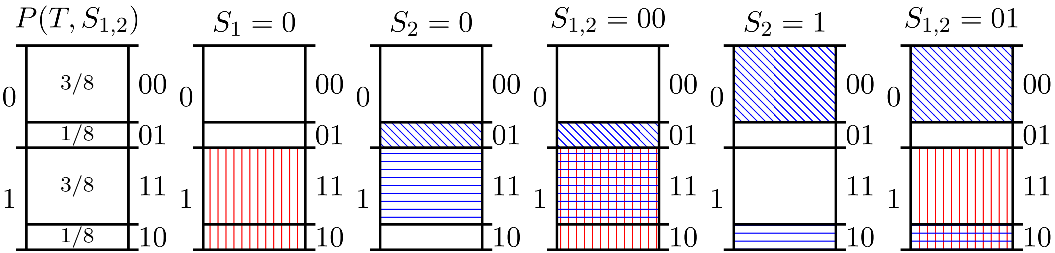
Robin AA Ince. Measuring multivariate redundant information with pointwise common change in surprisal. *Entropy*, 19 (7):318, 2017.

Claude E Shannon. A mathematical theory of communication. *Bell Syst. Tech. J.*, 27:623–656, 1948.

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

Philip M Woodward. *Probability and information theory: with applications to radar*. Pergamon, 1953.

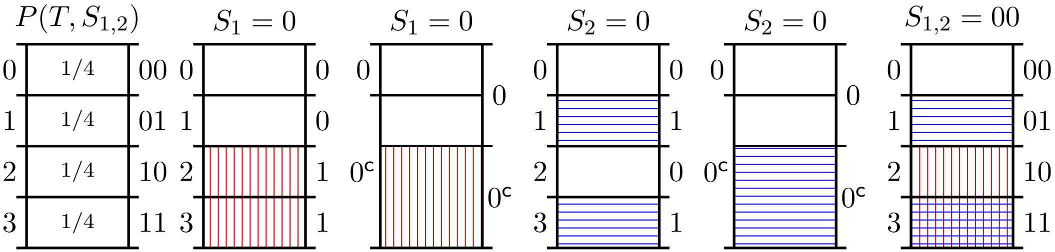
Example: IMPRDN



p	s_1	s_2	t	i_1^+	i_1^-	i_2^+	i_2^-	i_{12}^+	$i_{1,2}^-$	r^+	u_1^+	u_2^+	c^+	r^-	u_1^-	u_2^-	c^-
3/8	0	0	0	1	0	1	$\lg 4/3$	$\lg 8/3$	$\lg 4/3$	1	0	0	$\lg 4/3$	0	0	$\lg 4/3$	0
3/8	1	1	1	1	0	1	$\lg 4/3$	$\lg 8/3$	$\lg 4/3$	1	0	0	$\lg 4/3$	0	0	$\lg 4/3$	0
1/8	0	1	0	1	0	1	2	3	2	1	0	0	2	0	0	2	0
1/8	1	0	1	1	0	1	2	3	2	1	0	0	2	0	0	2	0
Expected values				1	0	1	0.811	1.811	0.811	1	0	0	0.811	0	0	0.811	0

$R(S_1, S_2 \to T) = 1 \quad U(S_1 \setminus S_2 \to T) = 0 \quad U(S_2 \setminus S_1 \to T) = -0.811 \quad C(S_1, S_2 \to T) = 0.811$

Example: TwoBitCopy



p	s_1	s_2	t	$t_{1,2}$	$t_{1,3}$	$t_{2,3}$	i_1^+	i_1^-	i_2^+	i_2^-	i_{12}^+	i_{12}^-	r^+	u_1^+	u_2^+	c^+	r^-	u_1^-	u_2^-	c^-
1/4	0	0	0	00	00	00	1	0	1	0	2	0	1	0	0	1	0	0	0	0
1/4	0	1	1	01	01	11	1	0	1	0	2	0	1	0	0	1	0	0	0	0
1/4	1	0	2	10	11	01	1	0	1	0	2	0	1	0	0	1	0	0	0	0
1/4	1	1	3	11	10	10	1	0	1	0	2	0	1	0	0	1	0	0	0	0
Expected values							1	0	1	0	2	0	1	0	0	1	0	0	0	0

$R(S_1, S_2 \to T) = 1$ $\quad U(S_1 \setminus S_2 \to T) = 0$ $\quad U(S_2 \setminus S_1 \to T) = 0$ $\quad C(S_1, S_2 \to T) = 1$