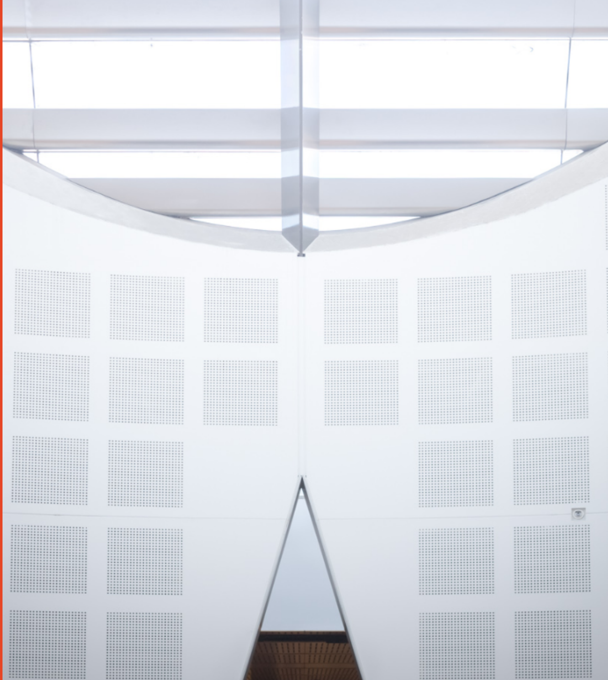
Pointwise Partial Information Decomposition Using Specificity and Ambiguity Lattices

UC Davis

Conor Finn

September 14, 2017



Pointwise Information Theory

- (Shannon) entropy: expected information in a realisation of a random variable

$$H(X) = \sum_x p(x) \log \tfrac{1}{p(x)} \geq 0$$

- Mutual information (MI) quantifies the expected information provided by Y about X, or vice versa

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

- From four postulates, Fano (1961) derived the **pointwise** mutual information which quantifies the information provided by event y about the event x, or vice versa

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

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

Pointwise mutual information can be negative!

Unique, Redundant and Synergistic Information

Consider three random variables S_1, S_2 and T and suppose we are interested in predicting the value of T from S_1 and S_2

- **Unique information** $U(T : S_1 \setminus S_2)$
  Source S_1 may contain information about T that source S_2 does not (or vice versa)

p	s_1	s_2	t
¼	0	0	0
¼	0	1	0
¼	1	0	1
¼	1	1	1

- **Redundant information** $R(T : S_1, S_2)$
  Source S_2 may contain the same information as source S_2 about T

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

- **Synergistic information** $C(T : S_1, S_2)$
  It is possible that neither source Z nor source Y contain information about X, yet take together they do; e.g. XOR

t	s_1	s_2	t
¼	0	0	0
¼	0	1	1
¼	1	0	1
¼	1	1	0

Information Decomposition

In general, unique, redundant and synergistic information are present simultaneously

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

- Mutual information captures

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

- Joint mutual information captures

$$I(T;S_1S_2) = R(T:S_1,S_2) + U(T:S_1\backslash S_2) + U(T:S_2\backslash S_1) + C(T:S_1,S_2)$$

- Three equations with four unknowns
  - Define one of the unique, redundant, or synergistic information and solve

- We would like to generalise these notions to any number of variables

Partial Information Decomposition

- PID of Williams and Beer (2010) provides an axiomatic framework for extending information decomposition to arbitrary number of source variables

Axioms (PID)

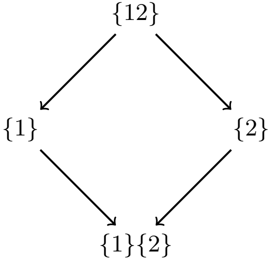
(1) *Symmetry*: $R(T : S_1, \ldots, S_n)$ is invariant under permutations of the S_i's

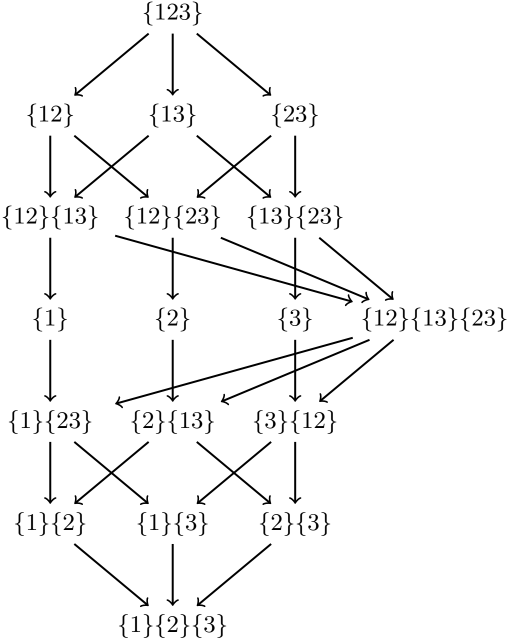
(2) *Self-redundancy*: $R(T : S_i) = I(T; S_i)$

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

- Based upon the idea that redundancy is in some way analogous to set intersection
- Yields a structure for multivariate information called the **redundancy lattice**
- No well-accepted definition of unique, redundant, or synergistic information which is compatible with PID in general has emerged

Redundancy lattice





Pointwise Information Decomposition

- In principle, we should be able to decompose pointwise information for each realisation

$$
\begin{aligned}
i(t;s_1) &= r(t:s_1,s_2) + u(t:s_1\backslash s_2)\\
i(t;s_2) &= r(t:s_1,s_2) + u(t:s_2\backslash s_1)\\
i(t;s_1 s_2) &= r(t:s_1,s_2) + u(t:s_1\backslash s_2) + u(t:s_2\backslash s_1) + c(t:s_1,s_2)
\end{aligned}
$$

Pointwise Information Decomposition

- In principle, we should be able to decompose pointwise information for each realisation

$$\begin{aligned} i(t;s_1) &= r(t:s_1,s_2) + u(t:s_1\backslash s_2) \\ i(t;s_2) &= r(t:s_1,s_2) + u(t:s_2\backslash s_1) \\ i(t;s_1s_2) &= r(t:s_1,s_2) + u(t:s_1\backslash s_2) + u(t:s_2\backslash s_1) + c(t:s_1,s_2) \end{aligned}$$

- These should average to yield the (average) information decomposition

$$\begin{aligned} R(T:S_1,S_2) &= \big\langle r(t:s_1,s_2)\big\rangle & \qquad U(T:S_1\backslash S_2) &= \big\langle u(t:s_1\backslash s_2)\big\rangle \\ C(T:S_1,S_2) &= \big\langle c(t:s_1,s_2)\big\rangle & \qquad U(T:S_2\backslash S_1) &= \big\langle u(t:s_2\backslash s_1)\big\rangle \end{aligned}$$

Pointwise Information Decomposition

- In principle, we should be able to decompose pointwise information for each realisation

$$
\begin{aligned}
i(t; s_1) &= r(t : s_1, s_2) + u(t : s_1 \backslash s_2) \\
i(t; s_2) &= r(t : s_1, s_2) + u(t : s_2 \backslash s_1) \\
i(t; s_1 s_2) &= r(t : s_1, s_2) + u(t : s_1 \backslash s_2) + u(t : s_2 \backslash s_1) + c(t : s_1, s_2)
\end{aligned}
$$

- These should average to yield the (average) information decomposition

$$
\begin{aligned}
R(T : S_1, S_2) &= \langle r(t : s_1, s_2) \rangle & \qquad U(T : S_1 \backslash S_2) &= \langle u(t : s_1 \backslash s_2) \rangle \\
C(T : S_1, S_2) &= \langle c(t : s_1, s_2) \rangle & \qquad U(T : S_2 \backslash S_1) &= \langle u(t : s_2 \backslash s_1) \rangle
\end{aligned}
$$

- And we should have the usual PID for the (average) informations

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

Example: PwUnq

- Consider the example called Pointwise Unique (PwUnq) from Finn et al. (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				

Example: PwUnq

- Consider the example called Pointwise Unique (PwUnq) from Finn et al. (2017b)

p	s_1	s_2	t	$i(t; s_1)$	$i(t; s_2)$	$i(t; s_1 s_2)$	
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	

Example: PwUnq

- Consider the example called Pointwise Unique (PwUnq) from Finn et al. (2017b)

p	s_1	s_2	t	$i(t; s_1)$	$i(t; s_2)$	$i(t; s_1 s_2)$	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

Example: PwUnq

- Consider the example called Pointwise Unique (PwUnq) from Finn et al. (2017b)

p	s_1	s_2	t	$i(t; s_1)$	$i(t; s_2)$	$i(t; s_1 s_2)$	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

Example: PwUnq

- Consider the example called Pointwise Unique (PwUnq) from Finn et al. (2017b)

p	s_1	s_2	t	$i(t; s_1)$	$i(t; s_2)$	$i(t; s_1 s_2)$	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 Williams and Beer (2010), Bertschinger et al. (2014), Griffith and Koch (2014) and Harder et al. (2013)

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

Pointwise Partial Information Decomposition

- Idea: rewrite PID axioms using pointwise MI instead of (average) MI

Axioms (PPID)

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

(2) *Self-redundancy*: $r(t : s_i) = i(t; s_i)$

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

Pointwise Partial Information Decomposition

- Idea: rewrite PID axioms using pointwise MI instead of (average) MI

Axioms (PPID)

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

(2) *Self-redundancy*: $r(t : s_i) = i(t; s_i)$

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

Problem: pointwise mutual information is not non-negative

- How do we deal with this issue?

Specificity and Ambiguity

- Idea: split the pointwise mutual information into non-negative entropic components

Specificity and Ambiguity

- Idea: split the pointwise mutual information into non-negative entropic components
- In Finn et al. (2017a), we proved pointwise information provided by s about t must be split in the following way

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

where

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

Specificity and Ambiguity

- Idea: split the pointwise mutual information into non-negative entropic components
- In Finn et al. (2017a), we proved pointwise information provided by s about t must be split in the following way

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

where

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

Axioms (PPID using Specificity and Ambiguity)

(1) *Symmetry*: $r^{\pm}(t : s_1, \ldots, s_n)$ is invariant under permutations of the s_i's
(2) *Self-redundancy*: $r^{\pm}(t : s_i) = i^{\pm}(t; s_i)$
(3) *Monotonicity*: $r^{\pm}(t : s_1, \ldots, s_n) \leq r^{\pm}(t : s_1; \ldots; s_{n-1})$

- Yields two lattices (per realisation): the specificity and ambiguity lattices

Redundancy Measure on the Lattices

- Still need a measure of redundant information on each lattice

Redundancy Measure on the Lattices

- Still need a measure of redundant information on each lattice
- We define the redundant specificity and redundant ambiguity to be

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

Redundancy Measure on the Lattices

- Still need a measure of redundant information on each lattice
- We define the redundant specificity and redundant ambiguity to be

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

- There is another axiom (Axiom 4) which helps justify this definition
  - Operational justification in terms of Kelly gambling

Example: PwUnq

p	s_1	s_2	t	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	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

Example: PwUnq

p	s_1	s_2	t	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	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

- Recombining the average specificities and average ambiguities yields the PID

$$R(T : S_1, S_2) = 1 - 1 = 0 \text{ bit} \qquad U(T : S_1 \backslash S_2) = 1/2 - 0 = 1/2 \text{ bit}$$

$$C(T : S_1, S_2) = 0 - 0 = 0 \text{ bit} \qquad U(T : S_2 \backslash S_1) = 1/2 - 0 = 1/2 \text{ bit}$$

- Matches the PPID suggested earlier

Example: XOR

p	s_1	s_2	t	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	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

- Recombining the average specificities and average ambiguities yields the PID

$$R(T : S_1, S_2) = 1 - 1 = 0 \text{ bit} \qquad U(T : S_1 \setminus S_2) = 0 - 0 = 0 \text{ bit}$$

$$C(T : S_1, S_2) = 1 - 0 = 1 \text{ bit} \qquad U(T : S_2 \setminus S_1) = 0 - 0 = 0 \text{ bit}$$

Example: XOR

p	s_1	s_2	t	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	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

- Recombining the average specificities and average ambiguities yields the PID

$$R(T : S_1, S_2) = 1 - 1 = 0 \text{ bit} \qquad U(T : S_1 \backslash S_2) = 0 - 0 = 0 \text{ bit}$$

$$C(T : S_1, S_2) = 1 - 0 = 1 \text{ bit} \qquad U(T : S_2 \backslash S_1) = 0 - 0 = 0 \text{ bit}$$

- Identifies redundancy due to shared knowledge from Bertschinger et al. (2013)

Example: IMPRDN

p	s_1	s_2	t	i_1^+	i_1^-	i_2^+	i_2^-	i_{12}^+	i_{12}^-	r^+	u_1^+	u_2^+	c^+	r^-	u_1^-	u_2^-	c^-
½	0	0	0	1	0	$\lg 8/5$	0	1	0	$\lg 8/5$	$\lg 5/4$	0	0	0	0	0	0
⅜	1	1	1	1	0	$\lg 8/3$	$\lg 4/3$	$\lg 8/3$	$\lg 4/3$	1	0	$\lg 4/3$	0	0	0	$\lg 4/3$	0
⅛	1	0	1	1	0	$\lg 8/5$	2	3	2	$\lg 8/5$	$\lg 5/4$	0	2	0	0	2	0
Expected				1	0	0.954	0.406	1.406	0.406	0.799	0.201	0.156	0.250	0	0	0.406	0

- Recombining the average specificities and average ambiguities yields the PID

$R(T : S_1, S_2) = 0.799 - 0 = 0.799$ bit $\quad U(T : S_1 \backslash S_2) = 0.201 - 0 = 0.201$ bit

$C(T : S_1, S_2) = 0.25 - 0.25 = 1$ bit $\quad U(T : S_2 \backslash S_1) = 0.156 - 0.406 = -0.25$ bit

Example: ImpRdn

p	s_1	s_2	t	i_1^+	i_1^-	i_2^+	i_2^-	i_{12}^+	i_{12}^-	r^+	u_1^+	u_2^+	c^+	r^-	u_1^-	u_2^-	c^-
1/2	0	0	0	1	0	$\lg 8/5$	0	1	0	$\lg 8/5$	$\lg 5/4$	0	0	0	0	0	0
3/8	1	1	1	1	0	$\lg 8/3$	$\lg 4/3$	$\lg 8/3$	$\lg 4/3$	1	0	$\lg 4/3$	0	0	0	$\lg 4/3$	0
1/8	1	0	1	1	0	$\lg 8/5$	2	3	2	$\lg 8/5$	$\lg 5/4$	0	2	0	0	2	0
Expected				1	0	0.954	0.406	1.406	0.406	0.799	0.201	0.156	0.250	0	0	0.406	0

- Recombining the average specificities and average ambiguities yields the PID

$$R(T : S_1, S_2) = 0.799 - 0 = 0.799 \text{ bit} \qquad U(T : S_1 \backslash S_2) = 0.201 - 0 = 0.201 \text{ bit}$$

$$C(T : S_1, S_2) = 0.25 - 0.25 = 1 \text{ bit} \qquad U(T : S_2 \backslash S_1) = 0.156 - 0.406 = -0.25 \text{ bit}$$

- May be negative unique information on average if a source is uniquely misinformative

Comparison to Other Decompositions and Measures

- Approach is most similar to Ince (2017) but differs in how the non-positivity is dealt with

Comparison to Other Decompositions and Measures

- Approach is most similar to Ince (2017) but differs in how the non-positivity is dealt with
- Similar to I_{min} of Williams and Beer (2010) but now fully pointwise

Comparison to Other Decompositions and Measures

- Approach is most similar to Ince (2017) but differs in how the non-positivity is dealt with
- 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 $\widetilde{UI}$

Comparison to Other Decompositions and Measures

- Approach is most similar to Ince (2017) but differs in how the non-positivity is dealt with
- 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)

Comparison to Other Decompositions and Measures

- Approach is most similar to Ince (2017) but differs in how the non-positivity is dealt with
- 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)
- Like other measures, there is no target monotonicity, i.e. don not have that

$$R_{\mathsf{min}}\big(S_1, S_2 \to T_1\big) \leq R_{\mathsf{min}}\big(S_1, S_2 \to T_1, T_2\big)$$

Comparison to Other Decompositions and Measures

- Approach is most similar to Ince (2017) but differs in how the non-positivity is dealt with
- 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)
- Like other measures, there is no target monotonicity, i.e. don not have that
$$R_{\text{min}}\big(S_1, S_2 \to T_1\big) \leq R_{\text{min}}\big(S_1, S_2 \to T_1, T_2\big)$$
- But unlike other measures, there is a target chain rule
$$R_{\text{min}}\big(S_1, S_2 \to T_1, T_2\big) = R_{\text{min}}\big(S_1, S_2 \to T_1\big) + R_{\text{min}}\big(S_1, S_2 \to T_2 | T_1\big)$$

Example: TwoBitCopy

p	s_1	s_2	t	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	00	1	0	1	0	2	0	1	0	0	1	0	0	0	0
1/4	0	1	01	1	0	1	0	2	0	1	0	0	1	0	0	0	0
1/4	1	0	10	1	0	1	0	2	0	1	0	0	1	0	0	0	0
1/4	1	1	11	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

- Recombining the average specificities and average ambiguities yields the PID

$$R(T : S_1, S_2) = 1 - 0 = 1 \text{ bit} \qquad U(T : S_1 \setminus S_2) = 0 - 0 = 0 \text{ bit}$$

$$C(T : S_1, S_2) = 1 - 0 = 1 \text{ bit} \qquad U(T : S_2 \setminus S_1) = 0 - 0 = 0 \text{ bit}$$

- Result is the same as it is for I_{min} of Williams and Beer (2010)

Example: TwoBitCopy

p	s_1	s_2	t	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	00	1	0	1	0	2	0	1	0	0	1	0	0	0	0
1/4	0	1	01	1	0	1	0	2	0	1	0	0	1	0	0	0	0
1/4	1	0	10	1	0	1	0	2	0	1	0	0	1	0	0	0	0
1/4	1	1	11	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

- Recombining the average specificities and average ambiguities yields the PID

$$R(T : S_1, S_2) = 1 - 0 = 1 \text{ bit} \qquad U(T : S_1 \setminus S_2) = 0 - 0 = 0 \text{ bit}$$

$$C(T : S_1, S_2) = 1 - 0 = 1 \text{ bit} \qquad U(T : S_2 \setminus S_1) = 0 - 0 = 0 \text{ bit}$$

- Result is the same as it is for I_{min} of Williams and Beer (2010)
- The measure and decomposition does not possess the identity property
  - Does mean that we can use this decomposition for more than 3 variables

Example: TwoBitCopy Horse Race

p	s_1	s_2	t	i_1^+	i_1^-	i_2^+	i_2^-	i_{12}^+	i_{12}^-	r^+	u_1^+	u_2^+	c^+	r^-	u_1^-	u_2^-	c^-
¼	b	r	a	1	0	1	0	2	0	1	0	0	1	0	0	0	0
¼	b	g	b	1	0	1	0	2	0	1	0	0	1	0	0	0	0
¼	w	r	c	1	0	1	0	2	0	1	0	0	1	0	0	0	0
¼	w	g	d	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

- Recombining the average specificities and average ambiguities yields the PID

$$R(T : S_1, S_2) = 1 - 0 = 1 \text{ bit} \qquad U(T : S_1 \backslash S_2) = 0 - 0 = 0 \text{ bit}$$

$$C(T : S_1, S_2) = 1 - 0 = 1 \text{ bit} \qquad U(T : S_2 \backslash S_1) = 0 - 0 = 0 \text{ bit}$$

- Result is the same as it is for I_{min} of Williams and Beer (2010)
- The measure and decomposition does not possess the identity property
  - Does mean that we can use this decomposition for more than 3 variables

References

Nils Bertschinger, Johannes Rauh, Eckehard Olbrich, and Jürgen Jost. Shared informationnew insights and problems in decomposing information in complex systems. In *Proceedings of the European Conference on Complex Systems 2012*, pages 251–269. Springer, 2013.

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, Mikhail Prokopenko, and Joseph T. Lizier. Decomposing pointwise information into directed positive and negative components. 2017a. To appear.

Conor Finn, Mikhail Prokopenko, 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.