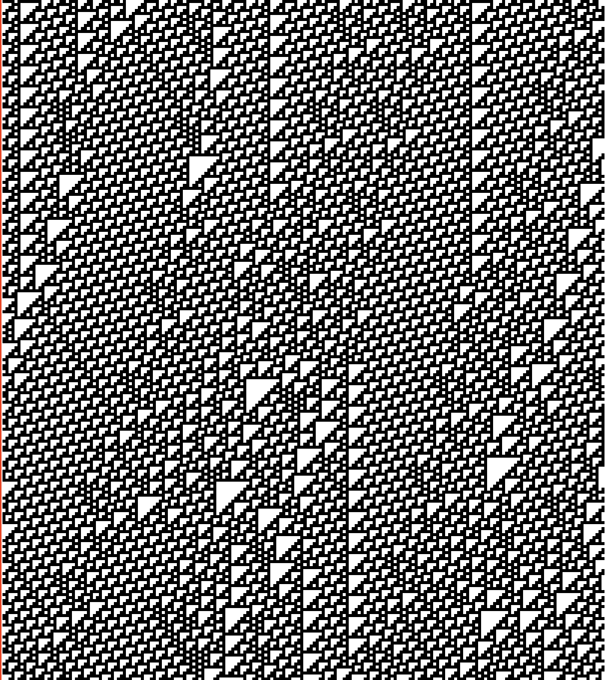
Generalised Measures of Multivariate Information Content

Conor Finn

March 2020

THE UNIVERSITY OF SYDNEY



Fundamentals of information theory

- The information content of an event

$$h(x) = \log \frac{1}{p(x)}$$

Fundamentals of information theory

- The information content of an event

$$h(x) = \log \frac{1}{p(x)}$$

  – the less likely the event, the most surprising it is $\implies 1/p(x)$

  – if two events are independent, then we should have

$$h(x,y) = h(x) + h(y) \iff p(x,y) = p(x) \cdot p(y) \implies \log$$

- Duality between information and surprise

Fundamentals of information theory

- The information content of an event

$$h(x) = \log \frac{1}{p(x)}$$

  - the less likely the event, the most surprising it is $\implies 1/p(x)$
  - if two events are independent, then we should have

$$h(x,y) = h(x) + h(y) \iff p(x,y) = p(x) \cdot p(y) \implies \log$$

- Duality between information and surprise
- The entropy is the average information content of a variable

$$H(X) = \mathrm{E}_X\left[h(x)\right] = \sum_{x \in \mathcal{X}} -p(x) \log p(x)$$

- Duality between average information and average surprise (uncertainty)

Conditional entropy and mutual information

- The entropy satisfies

$$H(X)+H(Y) \geq H(X,Y) \geq H(X),\ H(Y) \geq 0$$

- Conditional entropy

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

- Mutual information

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

Conditional entropy and mutual information

- The entropy satisfies

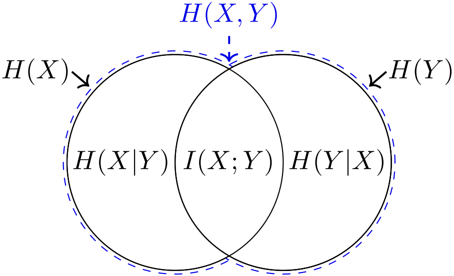
$$H(X)+H(Y) \geq H(X,Y) \geq H(X),\ H(Y) \geq 0$$

- Conditional entropy

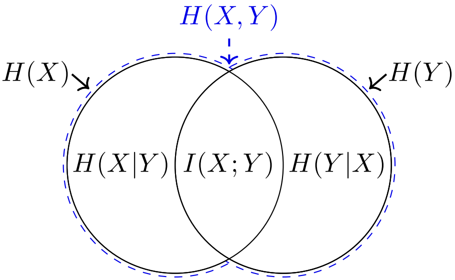
$$H(X|Y) = H(X,Y) - H(Y) \geq 0$$
$$H(Y|X) = H(X,Y) - H(X) \geq 0$$

- Mutual information

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



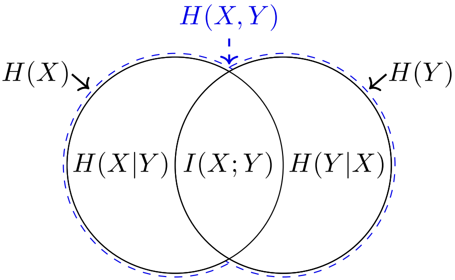
Analogy between entropy and measure

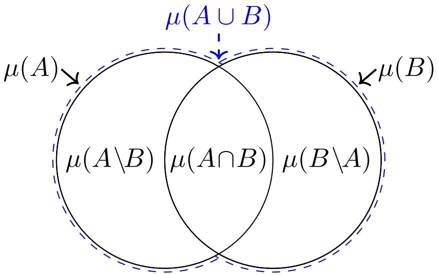


$$H(X) + H(Y) \geq H(X,Y) \geq H(X),\ H(Y) \geq 0$$

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

Analogy between entropy and measure





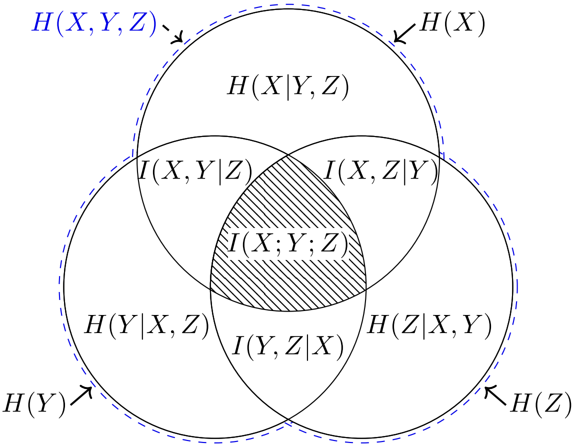
$$H(X) + H(Y) \geq H(X,Y) \geq H(X),\ H(Y) \geq 0$$

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

$$\mu(A) + \mu(B) \geq \mu(A \cup B) \geq \mu(A),\ \mu(B) \geq 0$$

$$\mu(A \backslash B) = \mu(A \cup B) - \mu(B) \geq 0$$
$$\mu(B \backslash A) = \mu(A \cup B) - \mu(A) \geq 0$$
$$\mu(A \cap B) = \mu(A) + \mu(B) - \mu(A \cup B) \geq 0$$

Analogy breaks down for three or more variables



- Multivariate mutual information (MMI)

$$
\begin{aligned}
I(X;Y;Z) = {} & H(X) + H(Y) + H(Z) \\
& - H(X,Y) - H(X,Z) - H(Y,Z) \\
& + H(Z,Y,Z)
\end{aligned}
$$

- MMI is not non-negative
- MMI has “no intuitive meaning”

Can we introduce information-theoretic measures that are analogous to measure for more than three variables?

Pointwise mutual information (PMI)

- The information content satisfies

$$h(x,y) \geq h(x),\, h(y) \geq 0$$

- Conditional information content

$$h(x|y) = h(x,y) - h(y) \geq 0$$
$$h(y|x) = h(x,y) - h(x) \geq 0$$

- However, we do **not** have the following

$$h(x) + h(y) \geq h(x,y) \geq 0$$

- Pointwise mutual information can be negative

$$i(x;y) = h(x) + h(y) - h(x,y)$$

Pointwise mutual information (PMI)

- The information content satisfies

$$h(x,y) \geq h(x),\, h(y) \geq 0$$

- Conditional information content

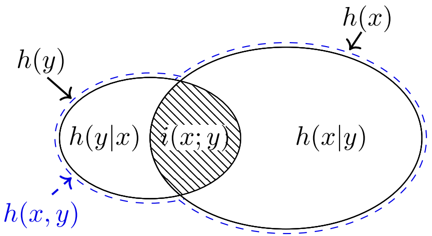
$$h(x|y) = h(x,y) - h(y) \geq 0$$
$$h(y|x) = h(x,y) - h(x) \geq 0$$

- However, we do **not** have the following

$$h(x) + h(y) \geq h(x,y) \geq 0$$

- Pointwise mutual information can be negative

$$i(x;y) = h(x) + h(y) - h(x,y)$$



Why can the pointwise mutual information be negative?

Why can the pointwise mutual information be negative?

	Johnny	Alice	Bob	Indy
				
Observations:	(X, Y)	X	Y	-
Knows:	$P(X, Y)$	$P(X)$	$P(Y)$	$P(X)$ & $P(Y)$

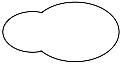
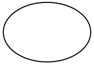
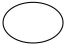
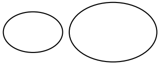
Why can the pointwise mutual information be negative?

	Johnny	Alice	Bob	Indy
				
Observations:	(X, Y)	X	Y	-
Knows:	$P(X, Y)$	$P(X)$	$P(Y)$	$P(X)$ & $P(Y)$
Realisation:	(x, y)	x	y	(x, y)

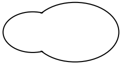
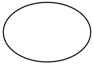
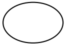
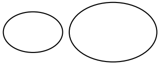
Why can the pointwise mutual information be negative?

	Johnny	Alice	Bob	Indy
				
Observations:	(X, Y)	X	Y	-
Knows:	$P(X, Y)$	$P(X)$	$P(Y)$	$P(X)$ & $P(Y)$
Realisation:	(x, y)	x	y	(x, y)
Information:	$h(x, y)$	$h(x)$	$h(y)$	$h(x) + h(y)$

Why can the pointwise mutual information be negative?

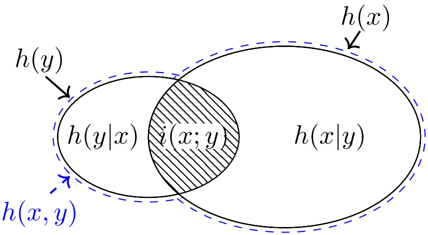
	Johnny	Alice	Bob	Indy
				
Observations:	(X, Y)	X	Y	-
Knows:	$P(X, Y)$	$P(X)$	$P(Y)$	$P(X)$ & $P(Y)$
Realisation:	(x, y)	x	y	(x, y)
Information:	$h(x, y)$	$h(x)$	$h(y)$	$h(x) + h(y)$
Venn diagram:				

Why can the pointwise mutual information be negative?

	Johnny	Alice	Bob	Indy
				
Observations:	(X, Y)	X	Y	-
Knows:	$P(X, Y)$	$P(X)$	$P(Y)$	$P(X)$ & $P(Y)$
Realisation:	(x, y)	x	y	(x, y)
Information:	$h(x, y)$	$h(x)$	$h(y)$	$h(x) + h(y)$
Venn diagram:				

$$h(x, y) \geq h(x),\ h(y) \geq 0$$
$$h(x|y) = h(x, y) - h(y) \geq 0$$
$$h(y|x) = h(x, y) - h(x) \geq 0$$

Venn diagram for information content

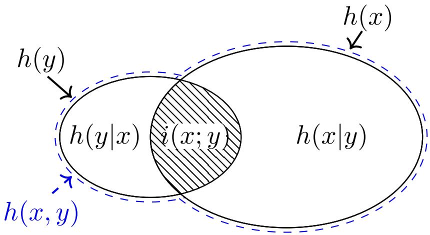


- Johnny has at least as much information as Alice and Bob

$$h(x,y) \geq h(x),\, h(y) \geq 0$$

$$h(x|y) = h(x,y) - h(y) \geq 0$$
$$h(y|x) = h(x,y) - h(x) \geq 0$$

Venn diagram for information content



- Johnny has at least as much information as Alice and Bob

$$h(x,y) \geq h(x),\; h(y) \geq 0$$

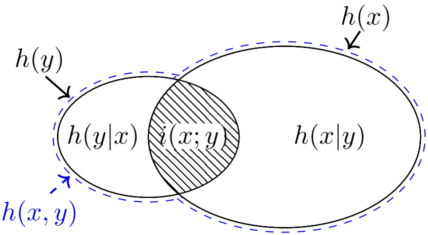
$$h(x|y) = h(x,y) - h(y) \geq 0$$
$$h(y|x) = h(x,y) - h(x) \geq 0$$

- The PMI can be negative because Indy can have more or less information than Johnny

$$i(x;y) = h(x) + h(y) - h(x,y)$$

- Sometimes Indy thinks he has more information than Johnny despite knowing less

Venn diagram for information content



- Johnny has at least as much information as Alice and Bob

$$h(x,y) \geq h(x),\ h(y) \geq 0$$

$$h(x|y) = h(x,y) - h(y) \geq 0$$
$$h(y|x) = h(x,y) - h(x) \geq 0$$

- The PMI can be negative because Indy can have more or less information than Johnny

$$i(x;y) = h(x) + h(y) - h(x,y)$$

- Sometimes Indy thinks he has more information than Johnny despite knowing less
- This happens because Indy **assumes** the marginal events are independent

- Sometimes Indy thinks he has more information than Johnny despite knowing less
- This happens because Indy **assumes** the marginal events are independent

- Sometimes Indy thinks he has more information than Johnny despite knowing less
- This happens because Indy **assumes** the marginal events are independent

Can we introduce a new information measure that is no less than Alice's and Bob's information, but is no greater than Johnny's information?

- Sometimes Indy thinks he has more information than Johnny despite knowing less
- This happens because Indy **assumes** the marginal events are independent

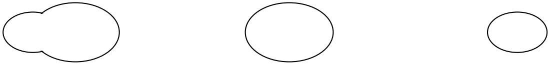
Can we introduce a new information measure that is no less than Alice's and Bob's information, but is no greater than Johnny's information?

- This would quantify the information associated with marginal information sharing

Marginal information sharing

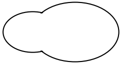
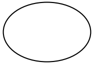
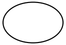
	Johnny	Alice	Bob	Eve
Observations:	(X, Y)	X	Y	-
Knows:	$P(X, Y)$	$P(X)$	$P(Y)$	$P(X)$ & $P(Y)$
Realisation:	(x, y)	x	y	(x, y)
Information:	$h(x, y)$	$h(x)$	$h(y)$	

Venn diagram:



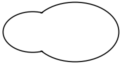
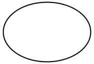
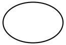
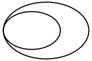
- Eve should have at least as much information as Alice and Bob, but no more than Johnny

Marginal information sharing

	Johnny	Alice	Bob	Eve
				
Observations:	(X, Y)	X	Y	-
Knows:	$P(X, Y)$	$P(X)$	$P(Y)$	$P(X)$ & $P(Y)$
Realisation:	(x, y)	x	y	(x, y)
Information:	$h(x, y)$	$h(x)$	$h(y)$	$\max\big(h(x), h(y)\big)$
Venn diagram:				

- Eve should have at least as much information as Alice and Bob, but no more than Johnny
- It is not difficult prove that Eve's information is given by $\max\big(h(x), h(y)\big)$

Marginal information sharing

	Johnny	Alice	Bob	Eve
				
Observations:	(X, Y)	X	Y	-
Knows:	$P(X, Y)$	$P(X)$	$P(Y)$	$P(X)$ & $P(Y)$
Realisation:	(x, y)	x	y	(x, y)
Information:	$h(x, y)$	$h(x)$	$h(y)$	$\max\big(h(x), h(y)\big)$
Venn diagram:				

- Eve should have at least as much information as Alice and Bob, but no more than Johnny
- It is not difficult prove that Eve's information is given by $\max\big(h(x), h(y)\big)$
- Eve's information will also be called the **union information** $h(x \sqcup y) = \max\big(h(x), h(y)\big)$

Union and intersection information content

- Union information content

$$h(x \sqcup y) = \max\big(h(x), h(y)\big)$$

satisfies

$$h(x)+h(y) \geq h(x \sqcup y) \geq h(x),\ h(y) \geq 0$$

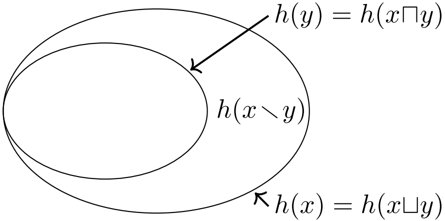
Union and intersection information content

- Union information content

$$h(x \sqcup y) = \max\big(h(x), h(y)\big)$$

satisfies

$$h(x) + h(y) \geq h(x \sqcup y) \geq h(x),\ h(y) \geq 0$$



Union and intersection information content

- Union information content

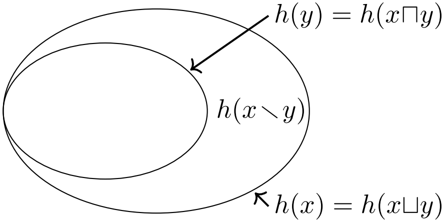
$$h(x \sqcup y) = \max\big(h(x), h(y)\big)$$

satisfies

$$h(x)+h(y) \geq h(x \sqcup y) \geq h(x),\ h(y) \geq 0$$

- Unique information content

$$\begin{aligned} h(x \setminus y) &= h(x \sqcup y) - h(y) \\ &= \max\big(h(x) - h(y), 0\big) \geq 0 \\ h(y \setminus x) &= h(x \sqcup y) - h(x) \\ &= \max\big(0, h(y) - h(x)\big) \geq 0 \end{aligned}$$



Union and intersection information content

- Union information content

$$h(x \sqcup y) = \max\big(h(x), h(y)\big)$$

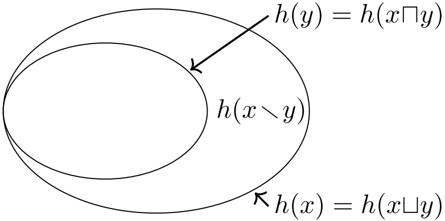
satisfies

$$h(x)+h(y) \geq h(x \sqcup y) \geq h(x),\ h(y) \geq 0$$

- Unique information content

$$h(x \setminus y) = h(x \sqcup y) - h(y) = \max\big(h(x) - h(y), 0\big) \geq 0$$

$$h(y \setminus x) = h(x \sqcup y) - h(x) = \max\big(0, h(y) - h(x)\big) \geq 0$$



- Intersection information content

$$h(x \sqcap y) = h(x) + h(y) - h(x \sqcup y) = \min\big(h(x), h(y)\big) \geq 0.$$

Union and intersection information content

- Union information content

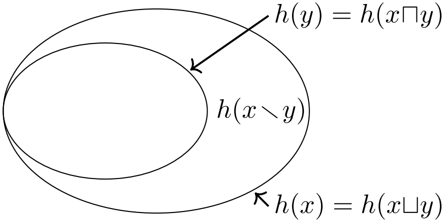
$$h(x \sqcup y) = \max\big(h(x), h(y)\big)$$

satisfies

$$h(x) + h(y) \geq h(x \sqcup y) \geq h(x),\ h(y) \geq 0$$

- Unique information content

$$\begin{aligned} h(x \setminus y) &= h(x \sqcup y) - h(y) \\ &= \max\big(h(x) - h(y), 0\big) \geq 0 \\ h(y \setminus x) &= h(x \sqcup y) - h(x) \\ &= \max\big(0, h(y) - h(x)\big) \geq 0 \end{aligned}$$



- Intersection information content

$$\begin{aligned} h(x \sqcap y) &= h(x) + h(y) - h(x \sqcup y) \\ &= \min\big(h(x), h(y)\big) \geq 0. \end{aligned}$$

- Decomposition

$$h(x \sqcup y) = h(x \sqcap y) + h(x \setminus y) + h(y \setminus x)$$

Union and intersection entropy

- Union entropy

$$H(X \sqcup Y) = \mathrm{E}_{XY}\left[h(x \sqcup y)\right]$$

satisfies

$$H(X)+H(Y) \geq H(X \sqcup Y) \geq H(X),\ H(Y) \geq 0$$

Union and intersection entropy

- Union entropy

$$H(X \sqcup Y) = \mathrm{E}_{XY}\left[h(x \sqcup y)\right]$$

satisfies

$$H(X)+H(Y) \geq H(X \sqcup Y) \geq H(X),\ H(Y) \geq 0$$

- Intersection information content

$$\begin{aligned} H(X \sqcap Y) &= H(X) + H(Y) - H(X \sqcup Y) \\ &= \mathrm{E}_{XY}\left[h(x \sqcap y)\right] \end{aligned}$$

Union and intersection entropy

- Union entropy

$$H(X \sqcup Y) = \mathrm{E}_{XY}\left[h(x \sqcup y)\right]$$

satisfies

$$H(X)+H(Y) \geq H(X \sqcup Y) \geq H(X),\ H(Y) \geq 0$$

- Unique information content

$$H(X \setminus Y) = \mathrm{E}_{XY}\left[h(x \setminus y)\right]$$
$$H(Y \setminus X) = \mathrm{E}_{XY}\left[h(y \setminus x)\right]$$

- Intersection information content

$$\begin{aligned} H(X \sqcap Y) &= H(X) + H(Y) - H(X \sqcup Y) \\ &= \mathrm{E}_{XY}\left[h(x \sqcap y)\right] \end{aligned}$$

Union and intersection entropy

- Union entropy

$$H(X \sqcup Y) = \mathrm{E}_{XY}\left[h(x \sqcup y)\right]$$

satisfies

$$H(X)+H(Y) \geq H(X \sqcup Y) \geq H(X),\ H(Y) \geq 0$$

- Unique information content

$$H(X \setminus Y) = \mathrm{E}_{XY}\left[h(x \smallsetminus y)\right]$$
$$H(Y \setminus X) = \mathrm{E}_{XY}\left[h(y \smallsetminus x)\right]$$

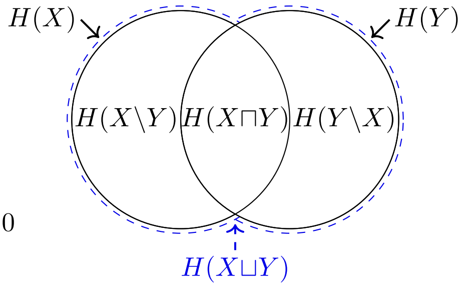
- Intersection information content

$$\begin{aligned} H(X \sqcap Y) &= H(X) + H(Y) - H(X \sqcup Y) \\ &= \mathrm{E}_{XY}\left[h(x \sqcap y)\right] \end{aligned}$$

- Decomposition

$$H(X \sqcup Y) = H(X \sqcap Y) + H(X \setminus Y) + H(Y \setminus X)$$

Union and intersection entropy



- Union entropy

$$H(X \sqcup Y) = \mathrm{E}_{XY}\left[h(x \sqcup y)\right]$$

satisfies

$$H(X) + H(Y) \geq H(X \sqcup Y) \geq H(X),\ H(Y) \geq 0$$

- Unique information content

$$H(X \setminus Y) = \mathrm{E}_{XY}\left[h(x \setminus y)\right]$$
$$H(Y \setminus X) = \mathrm{E}_{XY}\left[h(y \setminus x)\right]$$

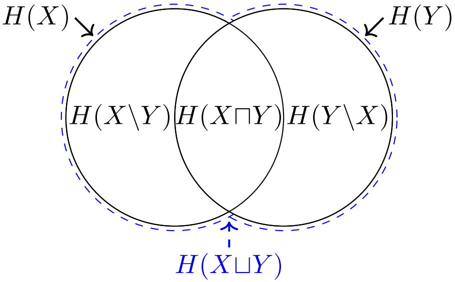
- Intersection information content

$$\begin{aligned} H(X \sqcap Y) &= H(X) + H(Y) - H(X \sqcup Y) \\ &= \mathrm{E}_{XY}\left[h(x \sqcap y)\right] \end{aligned}$$

- Decomposition

$$H(X \sqcup Y) = H(X \sqcap Y) + H(X \setminus Y) + H(Y \setminus X)$$

Analogy between shared marginal information and measure



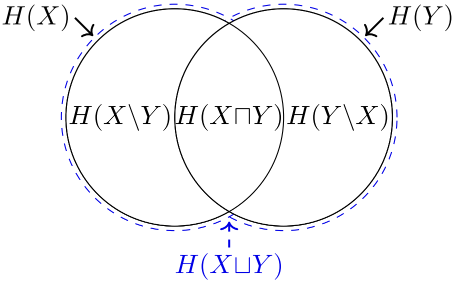
$$H(X) + H(Y) \geq H(X \sqcup Y) \geq H(X),\, H(Y) \geq 0$$

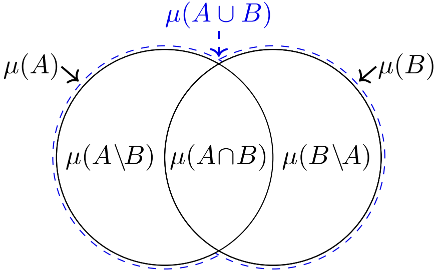
$$H(X\backslash Y) = H(X \sqcup Y) - H(Y) \geq 0$$

$$H(Y\backslash X) = H(X \sqcup Y) - H(X) \geq 0$$

$$H(X \sqcap Y) = H(X) + H(Y) - H(X \sqcup Y) \geq 0$$

Analogy between shared marginal information and measure





$$H(X)+H(Y)\geq H(X\sqcup Y)\geq H(X),\ H(Y)\geq 0$$

$$H(X\backslash Y)=H(X\sqcup Y)-H(Y)\geq 0$$

$$H(Y\backslash X)=H(X\sqcup Y)-H(X)\geq 0$$

$$H(X\sqcap Y)=H(X)+H(Y)-H(X\sqcup Y)\geq 0$$

$$\mu(A)+\mu(B)\geq\mu(A\cup B)\geq\mu(A),\ \mu(B)\geq 0$$

$$\mu(A\backslash B)=\mu(A\cup B)-\mu(B)\geq 0$$

$$\mu(B\backslash A)=\mu(A\cup B)-\mu(A)\geq 0$$

$$\mu(A\cap B)=\mu(A)+\mu(B)-\mu(A\cup B)\geq 0$$

Does this analogy hold for more than two variables?

Does this analogy hold for more than two variables?

- How do these new measures generalise?
- What happens when we share the marginal information through intermediaries?

Does this analogy hold for more than two variables?

- How do these new measures generalise?
- What happens when we share the marginal information through intermediaries?
- Consider the following:
  - Suppose that Alice and Bob share their information with Dan
  - The information Dan could have gotten from either is $h(x \sqcap y)$
  - Now say that Charlie and Dan share their information with Eve
  - Eve's information is now given by $h\big((x \sqcap y) \sqcup z\big)$

Does this analogy hold for more than two variables?

- How do these new measures generalise?
- What happens when we share the marginal information through intermediaries?
- Consider the following:
  - Suppose that Alice and Bob share their information with Dan
  - The information Dan could have gotten from either is $h(x \sqcap y)$
  - Now say that Charlie and Dan share their information with Eve
  - Eve's information is now given by $h\big((x \sqcap y) \sqcup z\big)$
- How many unique combinations or ways to share marginal information are there?

Algebraic structure

- Idempotent

$$h(x \sqcup x) = h(x)$$
$$h(x \sqcap x) = h(x)$$

Algebraic structure

- Idempotent

$$h(x \sqcup x) = h(x)$$
$$h(x \sqcap x) = h(x)$$

- Commutative

$$h(x \sqcup y) = h(y \sqcup x)$$
$$h(x \sqcap y) = h(y \sqcap x)$$

Algebraic structure

- Idempotent

$$h(x \sqcup x) = h(x)$$
$$h(x \sqcap x) = h(x)$$

- Commutative

$$h(x \sqcup y) = h(y \sqcup x)$$
$$h(x \sqcap y) = h(y \sqcap x)$$

- Associative

$$h(x \sqcup y \sqcup z) = h\big((x \sqcup y) \sqcup z\big) = h\big(x \sqcup (y \sqcup z)\big)$$
$$h(x \sqcap y \sqcap z) = h\big((x \sqcap y) \sqcap z\big) = h\big(x \sqcap (y \sqcap z)\big)$$

Algebraic structure

- Idempotent

$$h(x \sqcup x) = h(x)$$
$$h(x \sqcap x) = h(x)$$

- Commutative

$$h(x \sqcup y) = h(y \sqcup x)$$
$$h(x \sqcap y) = h(y \sqcap x)$$

- Associative

$$h(x \sqcup y \sqcup z) = h\big((x \sqcup y) \sqcup z\big) = h\big(x \sqcup (y \sqcup z)\big)$$
$$h(x \sqcap y \sqcap z) = h\big((x \sqcap y) \sqcap z\big) = h\big(x \sqcap (y \sqcap z)\big)$$

- Absorption

$$h\big(x \sqcup (x \sqcap y)\big) = h(x)$$
$$h\big(x \sqcap (x \sqcup y)\big) = h(x)$$

Algebraic structure

- Idempotent

$$h(x \sqcup x) = h(x)$$
$$h(x \sqcap x) = h(x)$$

- Commutative

$$h(x \sqcup y) = h(y \sqcup x)$$
$$h(x \sqcap y) = h(y \sqcap x)$$

- Associative

$$h(x \sqcup y \sqcup z) = h\big((x \sqcup y) \sqcup z\big) = h\big(x \sqcup (y \sqcup z)\big)$$
$$h(x \sqcap y \sqcap z) = h\big((x \sqcap y) \sqcap z\big) = h\big(x \sqcap (y \sqcap z)\big)$$

- Absorption

$$h\big(x \sqcup (x \sqcap y)\big) = h(x)$$
$$h\big(x \sqcap (x \sqcup y)\big) = h(x)$$

- Distributive

$$h\big(x \sqcup (y \sqcap z)\big) = h\big((x \sqcup y) \sqcap (x \sqcup z)\big)$$
$$h\big(x \sqcap (y \sqcup z)\big) = h\big((x \sqcap y) \sqcup (x \sqcap z)\big)$$

Algebraic structure

- Idempotent

$$h(x \sqcup x) = h(x)$$
$$h(x \sqcap x) = h(x)$$

- Absorption

$$h\big(x \sqcup (x \sqcap y)\big) = h(x)$$
$$h\big(x \sqcap (x \sqcup y)\big) = h(x)$$

- Commutative

$$h(x \sqcup y) = h(y \sqcup x)$$
$$h(x \sqcap y) = h(y \sqcap x)$$

- Distributive

$$h\big(x \sqcup (y \sqcap z)\big) = h\big((x \sqcup y) \sqcap (x \sqcup z)\big)$$
$$h\big(x \sqcap (y \sqcup z)\big) = h\big((x \sqcap y) \sqcup (x \sqcap z)\big)$$

- Associative

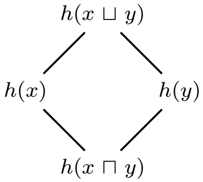
$$h(x \sqcup y \sqcup z) = h\big((x \sqcup y) \sqcup z\big) = h\big(x \sqcup (y \sqcup z)\big)$$
$$h(x \sqcap y \sqcap z) = h\big((x \sqcap y) \sqcap z\big) = h\big(x \sqcap (y \sqcap z)\big)$$

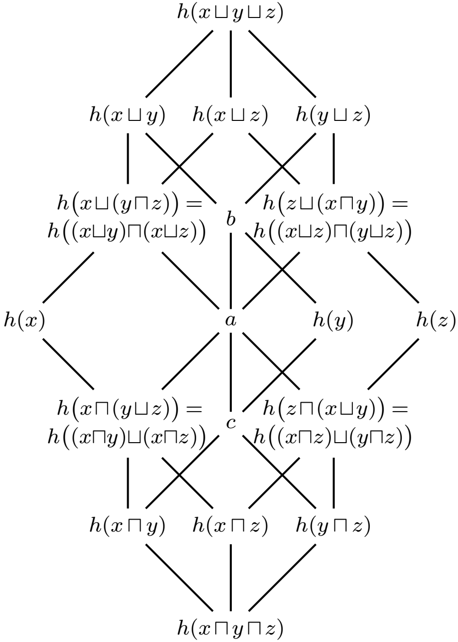
- These 5 properties yield a **distributive lattice**

$$
\begin{aligned}
a &= h\big((x \sqcup y) \sqcap (x \sqcup z) \sqcap (y \sqcup z)\big) \\
&= h\Big(\big(x \sqcup (y \sqcap z)\big) \sqcap \big(y \sqcup (x \sqcap z)\big)\Big) \\
&= h\Big(\big(x \sqcup (y \sqcap z)\big) \sqcap \big(z \sqcup (x \sqcap y)\big)\Big) \\
&= h\Big(\big(y \sqcup (x \sqcap z)\big) \sqcap \big(z \sqcup (x \sqcap y)\big)\Big) \\
&= h\Big(\big(y \sqcap (x \sqcup z)\big) \sqcup \big(z \sqcap (x \sqcup y)\big)\Big) \\
&= h\Big(\big(x \sqcap (y \sqcup z)\big) \sqcup \big(z \sqcap (x \sqcup y)\big)\Big) \\
&= h\Big(\big(x \sqcap (y \sqcup z)\big) \sqcup \big(y \sqcap (x \sqcup z)\big)\Big) \\
&= h\big((x \sqcap y) \sqcup (x \sqcap z) \sqcup (y \sqcap z)\big)
\end{aligned}
$$

$$b = h\big(y \sqcup (x \sqcap z)\big) = h\big((x \sqcup y) \sqcap (y \sqcup z)\big)$$

$$c = h\big(y \sqcap (x \sqcup z)\big) = h\big((x \sqcap y) \sqcup (y \sqcap z)\big)$$



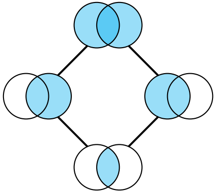


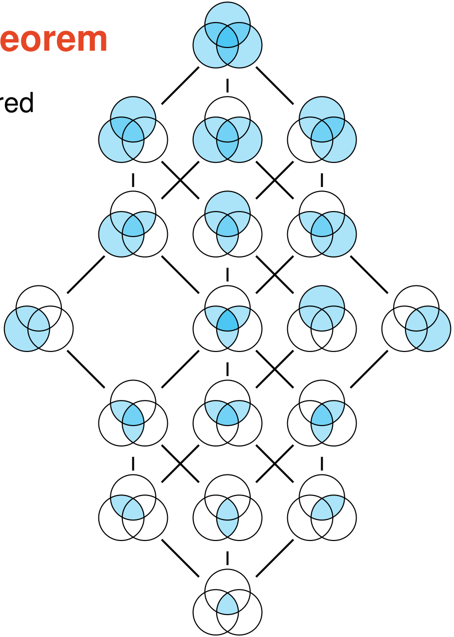
Birkhoff's representation theorem

- Equivalence relation between shared marginal information and sets

Birkhoff’s representation theorem

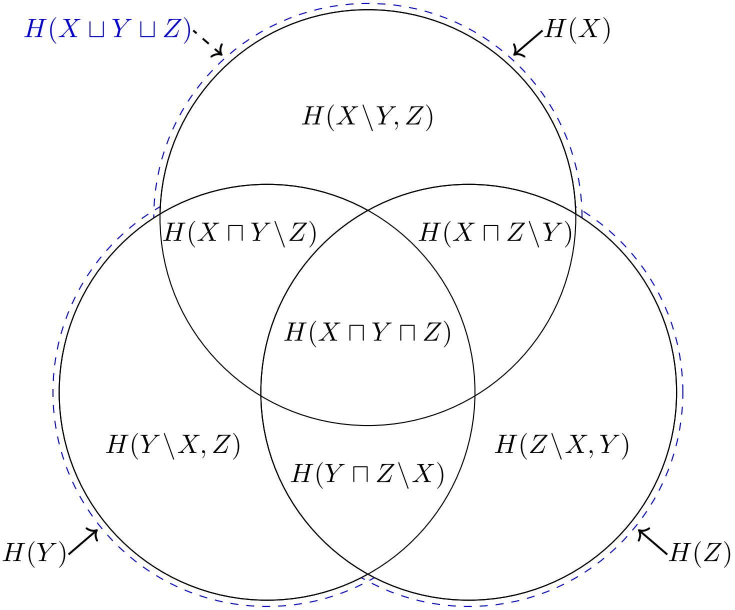
- Equivalence relation between shared marginal information and sets





Can we introduce information-theoretic measures that are analogous to measure for more than three variables?

Yes



$H(X \sqcup Y \sqcup Z)$
$H(X)$
$H(X \setminus Y, Z)$
$H(X \sqcap Y \setminus Z)$
$H(X \sqcap Z \setminus Y)$
$H(X \sqcap Y \sqcap Z)$
$H(Y \setminus X, Z)$
$H(Z \setminus X, Y)$
$H(Y \sqcap Z \setminus X)$
$H(Y)$
$H(Z)$

Union and intersection entropy are a new, fundamental contribution to information theory.

Union and intersection entropy are a new, fundamental contribution to information theory.

These measures can be utilised in information decomposition

Information decomposition

Consider trying to predict T from S_1 and S_2

- Several types of information

Information decomposition

Consider trying to predict T from S_1 and S_2

- Several types of information
  - **Unique information** $U(S_1 \setminus S_2 \rightarrow T)$

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

Information decomposition

Consider trying to predict T from S_1 and S_2

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

Information decomposition

Consider trying to predict T from S_1 and S_2

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

Information decomposition

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** $C(S_1, S_2 \to T)$

- Mutual information captures

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

Unq

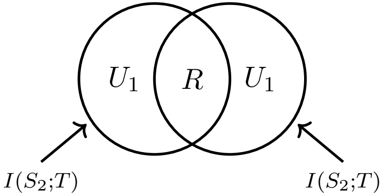
p	s_1	s_2	t
1/4	0	0	0
1/4	0	1	0
1/4	1	0	1
1/4	1	1	1

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



Information decomposition

Consider trying to predict T from S_1 and S_2

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

- Mutual information captures

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

- Joint mutual information captures

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

Unq

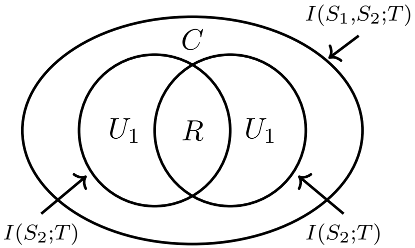
p	s_1	s_2	t
1/4	0	0	0
1/4	0	1	0
1/4	1	0	1
1/4	1	1	1

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



Partial information decomposition

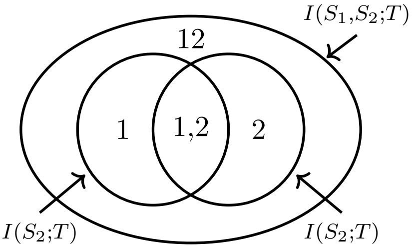
- Axioms for redundant information
  1. Commutativity
  2. Monotonically decreasing
  3. Self-redundancy (idempotency)

- Yields a **redundancy lattice**

Partial information decomposition

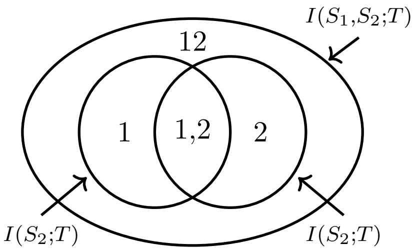
- Axioms for redundant information
  1. Commutativity
  2. Monotonically decreasing
  3. Self-redundancy (idempotency)

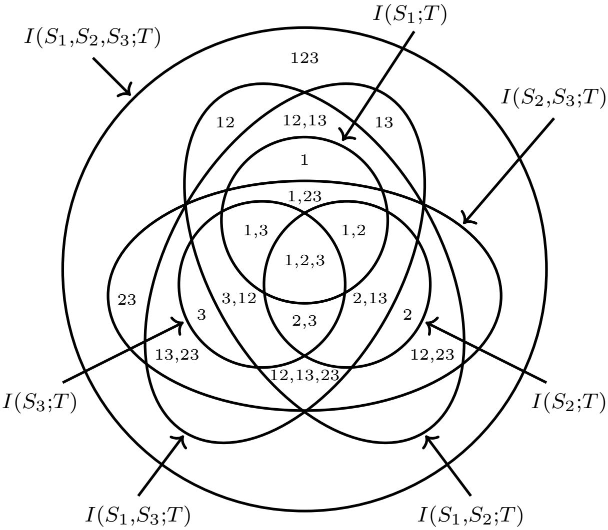
- Yields a **redundancy lattice**



Partial information decomposition

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





PID is elegant, however...

- Unique evaluation requires a definition of redundant information
  - providing this definition has been a contentious area of research
- Most appoaches do not work for two or more sources (not very useful)
- Information dynamics requires a pointwise information decomposition

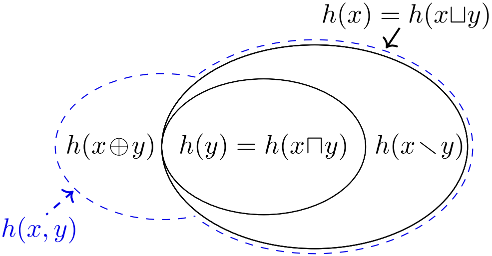
Use the intersection information content is a measure of redundancy

We can also interpret the synergistic information in this framework

Synergistic information content

- Eve has no more information than Johnny

$$h(x, y) \geq h(x \sqcup y)$$



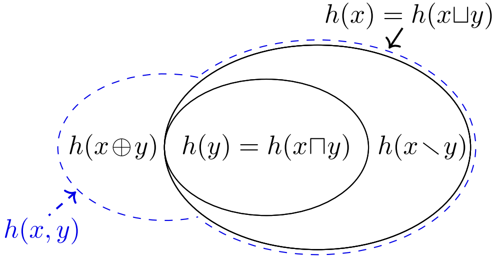
Synergistic information content

- Eve has no more information than Johnny

$$h(x, y) \geq h(x \sqcup y)$$

- Synergistic information content

$$h(x \oplus y) = h(x, y) - h(x \sqcup y)$$
$$= \min\left(h(y|x), h(x|y)\right) \geq 0$$



Synergistic information content

- Eve has no more information than Johnny

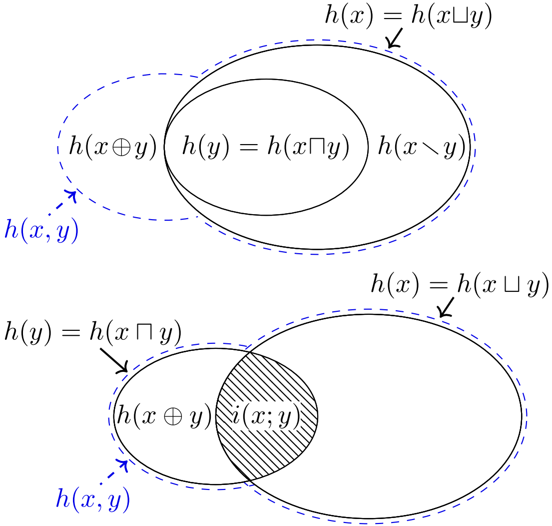
$$h(x,y) \geq h(x \sqcup y)$$

- Synergistic information content

$$h(x \oplus y) = h(x,y) - h(x \sqcup y)$$
$$= \min\big(h(y|x), h(x|y)\big) \geq 0$$

- Mutual information content

$$i(x;y) = h(x \sqcup y) - h(x \oplus y)$$



Synergistic information content

- Eve has no more information than Johnny

$$h(x,y) \geq h(x \sqcup y)$$

- Synergistic information content

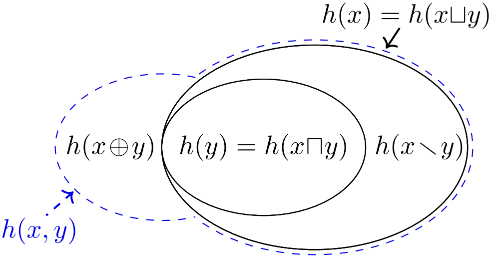
$$h(x \oplus y) = h(x,y) - h(x \sqcup y)$$
$$= \min\big(h(y|x), h(x|y)\big) \geq 0$$

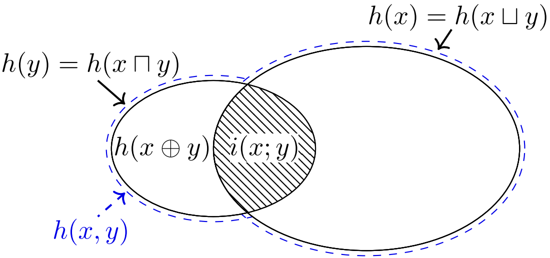
- Mutual information content

$$i(x;y) = h(x \sqcup y) - h(x \oplus y)$$

- Decomposition

$$h(x,y) = h(x \setminus y) + h(y \setminus x) + h(x \sqcap y) + h(x \oplus y)$$





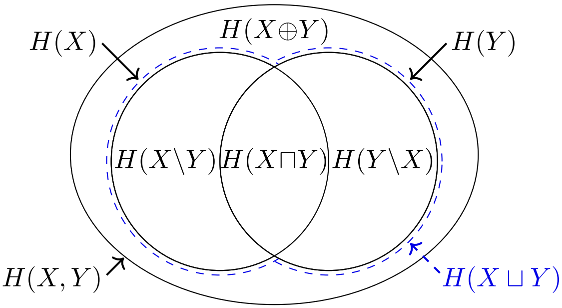
Synergistic entropy

- Synergistic entropy

$$H(X \oplus Y) = H(X,Y) - H(X \sqcup Y)$$
$$= \mathrm{E}_{XY}\big[h(x \oplus y))\big] \geq 0$$

- Mutual information

$$I(X;Y) = H(X \sqcup Y) - H(X \oplus Y)$$



- Decomposition

$$H(X,Y) = H(X \searrow Y) + H(Y \searrow X) + H(X \sqcap Y) + H(X \oplus Y)$$

We can also generalise the synergistic entropy to any number of variables

We can also generalise the synergistic entropy to any number of variables

- This is non-trivial
- It requires you to combine the algebraic structure of joint information with that of shared marginal information

We can also generalise the synergistic entropy to any number of variables

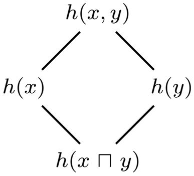
- This is non-trivial
- It requires you to combine the algebraic structure of joint information with that of shared marginal information
- It turns out the redundancy lattice is an substructure within the larger algebraic structure

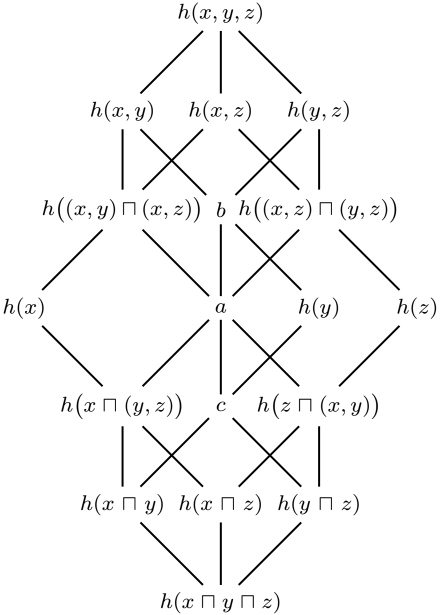
Redundancy lattice

$$a = h\big((x,y) \sqcap (x,z) \sqcap (y,z)\big)$$

$$b = h\big((x,y) \sqcap (y,z)\big)$$

$$c = h\big(y \sqcap (x,z)\big)$$

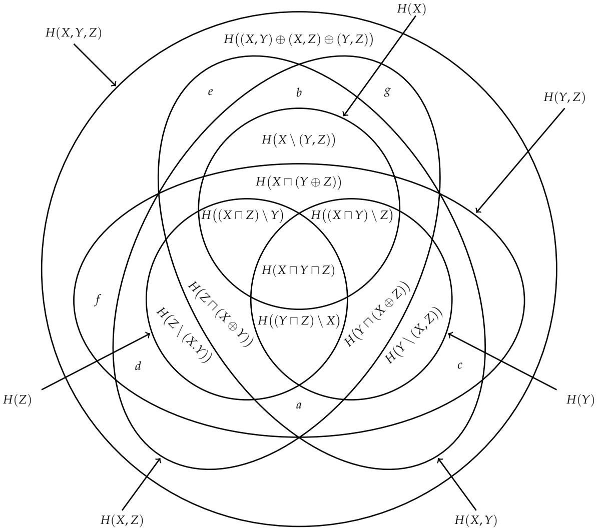




$$b = H\big((X \oplus Y) \sqcap (X \oplus Z) \setminus (Y,Z)\big) \qquad e = H\big((X \oplus Y) \setminus (X,Z) \sqcup (Y,Z)\big)$$

$$c = H\big((X \oplus Y) \sqcap (Y \oplus Z) \setminus (X,Z)\big) \qquad f = H\big((X \oplus Z) \setminus (X,Y) \sqcup (Y,Z)\big)$$

$$d = H\big((X \oplus Z) \sqcap (Y \oplus Z) \setminus (X,Y)\big) \qquad g = H\big((Y \oplus Z) \setminus (X,Y) \sqcup (X,Z)\big)$$



Conclusions

- The union and intersection entropy are new, fundamental contributions to information theory
- The are readily interpretable in terms of sets through an equivalence relation
- The redundancy lattice from partial information decomposition arises naturally within this framework

Thank you for listening

- Links to paper and copy of slides available at `https://finnconor.github.io/`

- C. Finn and J. T. Lizier. Generalised measures of multivariate information content. *Entropy*, 22(2):216, 2020.
- C. Finn and J. T. Lizier. Quantifying information modification in cellular automata using pointwise partial information decomposition. In *Artificial Life Conference Proceedings*, page 386. MIT Press, 2018.
- M. Wibral, C. Finn, P. Wollstadt, J. T. Lizier, and V. Priesemann. Quantifying information modification in developing neural networks via partial information decomposition. *Entropy*, 19(9):494, 2017.