Information Measures & Information Diagrams

Reading for this lecture:

CMR articles *Yeung & Anatomy*

Information Measures ...

One random variable: $X \sim \Pr(x)$

Use paradigm: One shot sampling.

Recall information theory quantity:

Entropy: $H[X]$

Information Measures ...

Two random variables: $X \sim \Pr(x)$

$$Y \sim \Pr(y)$$

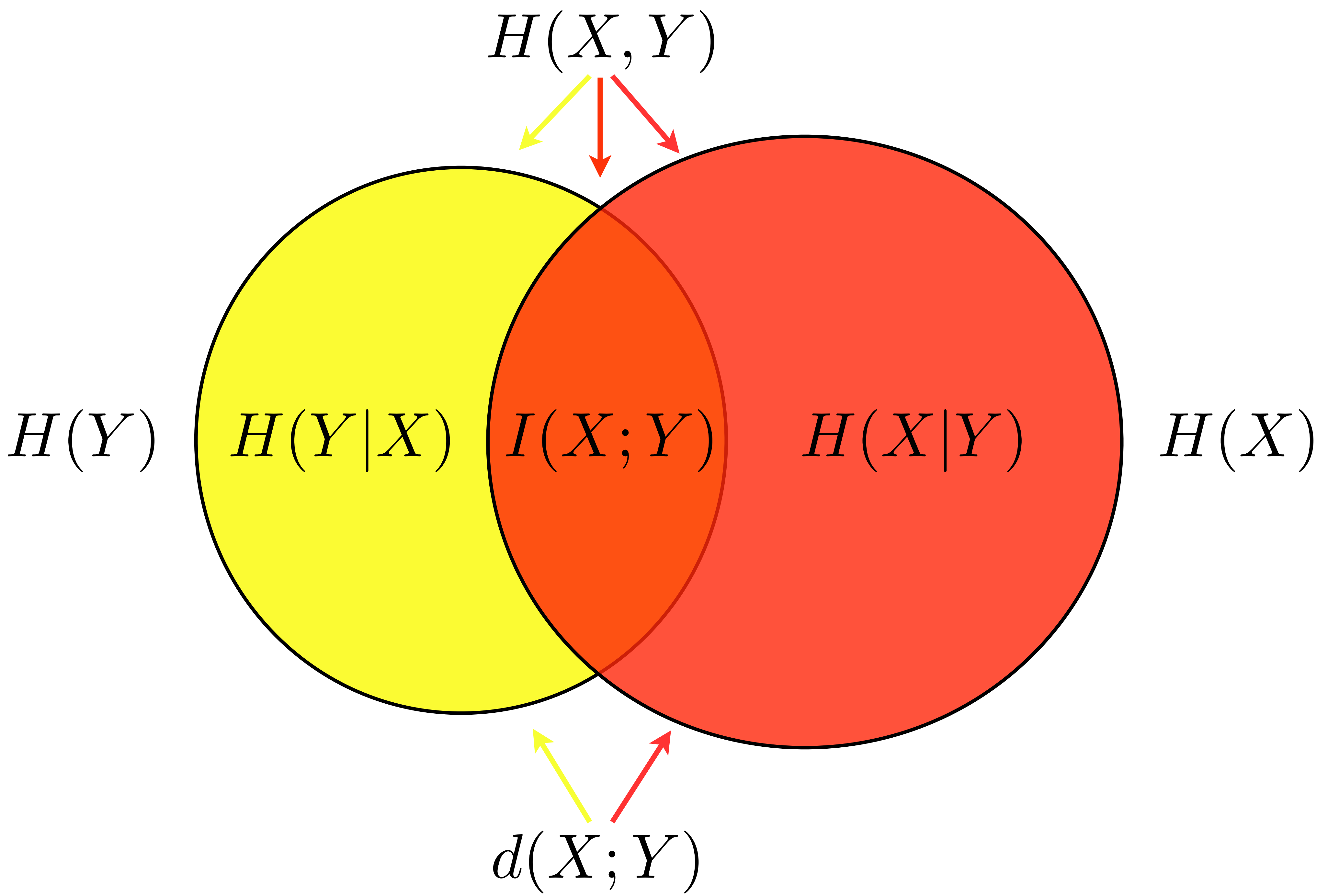
$$(X, Y) \sim \Pr(x, y)$$

Recall information theory quantities:

- Joint entropy: $H[X, Y]$
- Conditional Entropies: $H[X|Y] \quad H[Y|X]$
- Mutual Information: $I[X; Y]$
- Information Metric: $d(X, Y)$

Information Measures ...

Event Space Relationships of Information Quantifiers:



Information Measures ...

Just a cute mnemonic?

No, much more structure there.

Information Measures ...

Three random variables: $X \sim \Pr(x)$

$$Y \sim \Pr(y)$$

$$Z \sim \Pr(z)$$

$$(X, Y, Z) \sim \Pr(x, y, z)$$

Information theory quantities:

Joint entropy: $H[X, Y, Z]$

Conditional entropies: $H[X|Y, Z]$ $H[Y|X, Z]$ $H[Z|X, Y]$

$H[X, Y|Z]$ $H[X, Z|Y]$ $H[Y, Z|X]$

Conditional mutual information: $I[X; Y|Z]$

Mutual information? $I[X; Y; Z]$?

Information metric?

How to visualize?

A roadmap for information relationships?

Information Measures ...

N random variables?

$$(X_1, X_2, \ldots, X_N) \sim \Pr(x_1, x_2, \ldots, x_N)$$

Agenda:

Mechanics of information measures for N variables

Graphical information diagram for processes

Information Measures ...

The Shannon Information Measures:

Entropy, conditional entropy, mutual information,
and conditional mutual information

All information measures can be expressed as linear combination of entropies. E.g.,

$$H[X|Y] = H[X] - H[X,Y]$$

$$I[X;Y] = H[X] + H[Y] - H[X,Y]$$

$$d(X,Y) = H[X|Y] + H[Y|X]$$

The Shannon Information Measures ...

No fundamental difference btw entropy H & mutual information I.

In fact, entropy can be introduced as a "self-mutual information":

$$H[X] = I[X;X]$$

So, there's really a single quantity being referenced.

Information Measures ...

The Shannon Information Measures ...

I-Measure:

I and H form a *signed* measure over event space.

Set variables for event space:

$\widetilde{X}$ corresponds to random variable X

$\widetilde{Y}$ corresponds to random variable Y

Universal set: $\Omega = \widetilde{X} \cup \widetilde{Y}$

Over σ-field of atoms:

$$\mathcal{F} = \left\{ (\widetilde{X} \cup \widetilde{Y}), \widetilde{X}, \widetilde{Y}, (\widetilde{X} \cap \widetilde{Y}), (\widetilde{X} \cap \widetilde{Y}^c), (\widetilde{X}^c \cap \widetilde{Y}), (\widetilde{X} \cap \widetilde{Y})^c, \emptyset \right\}$$

Real-valued measure for atoms: $\mu^* \in \mathbb{R}$

The Shannon Information Measures ...

I-Measure ... defined

By the mappings:

$$\begin{aligned}
\mu^*(\widetilde{X} \cup \widetilde{Y}) &= H[X,Y] \\
\mu^*(\widetilde{X}) &= H[X] \\
\mu^*(\widetilde{Y}) &= H[Y] \\
\mu^*(\widetilde{X} \cap \widetilde{Y}) &= I[X;Y] \\
\mu^*(\widetilde{X} - \widetilde{Y}) &= H[X|Y] \\
\mu^*(\widetilde{Y} - \widetilde{X}) &= H[Y|X] \\
\mu^*((\widetilde{X} \cap \widetilde{Y})^c) &= H[X|Y] + H[Y|X] \\
\mu^*(\emptyset) &= 0
\end{aligned}$$

Where:

$$\begin{aligned}
\widetilde{X} - \widetilde{Y} &= \widetilde{X} \cap \widetilde{Y}^c \\
\widetilde{Y} - \widetilde{X} &= \widetilde{X}^c \cap \widetilde{Y}
\end{aligned}$$

Information Measures ...

The Shannon Information Measures:

I-Measure ...

Roadmap: Information measures to set-theoretic operations

$$\begin{array}{rcl} H \text{ and } I & \rightarrow & \mu^* \\ , & \rightarrow & \cup \\ ; & \rightarrow & \cap \\ | & \rightarrow & - \end{array}$$

Information Measures ...

The Shannon Information Measures:

I-Measure ...

For example,

$$I[X;Y] = H[X] + H[Y] - H[X,Y]$$

or

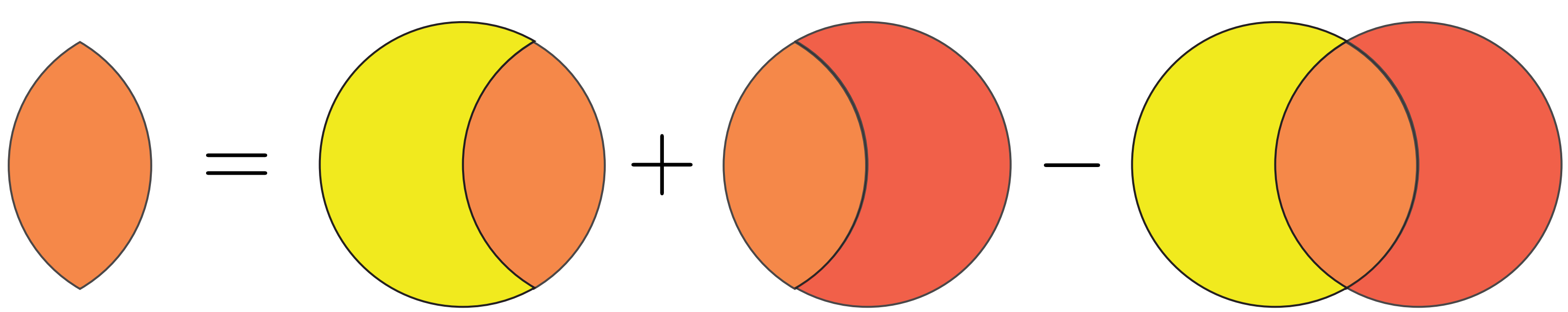
$$\mu^*(\widetilde{X} \cap \widetilde{Y}) = \mu^*(\widetilde{X}) + \mu^*(\widetilde{Y}) - \mu^*(\widetilde{X} \cup \widetilde{Y})$$

map to the set identity

$$\widetilde{X} \cap \widetilde{Y} = \widetilde{X} \cup \widetilde{Y} - (\widetilde{X} \cup \widetilde{Y})$$

(Notation ambiguous.)

That is,



Information Measures ...

The Shannon Information Measures:

I-Measure ...

Benefits:

- Extends to N variables
- Makes explicit underlying structure of information theory
- Graphical "calculus" of information identities
- Often easier to work with, sometimes not
- Often easier to discover new quantities:
  - e.g., N-variable mutual informations ... stay tuned

The Shannon Information Measures:

I-Measure for N variables:

Definition: $\mathcal{F}$ = σ-field generated by $W = \{X_i : i = 1, \ldots, N\}$

(Drop tildes on set variables!)

Universal set: $\Omega = \bigcup_i^N X_i$

Atoms: $A \in \mathcal{F} : A = \bigcap_{i=1}^N Y_i \ , \ Y_i = X_i \text{ or } X_i^c$

$$||A|| = 2^N - 1$$ (Max number of atoms)

$$||\mathcal{F}|| = 2^{2^N - 1}$$ (Max number of sets of atoms)

Information Measures ...

The Shannon Information Measures:

Theorem: I-Measure for N random variables:

$$G, G', G'' \in W = \{X_i, i = 1, \ldots N\}$$

$$\begin{aligned}
\mu^* \left(\bigcup_{X \in G} X \right) &\equiv H[X, X \in G] \\
\mu^* \left(\left(\bigcup_{X \in G} X \right) - \left(\bigcup_{Y \in G'} Y \right) \right) &\equiv H[X, X \in G | Y, Y \in G'] \\
\mu^* \left(\left(\bigcup_{X \in G} X \right) \bigcap \left(\bigcup_{Y \in G'} Y \right) \right) &\equiv I[X, X \in G; Y, Y \in G'] \\
\mu^* \left(\left(\bigcup_{X \in G} X \right) \bigcap \left(\bigcup_{Y \in G'} Y \right) - \left(\bigcup_{Z \in G''} Z \right) \right) &\equiv I[X, X \in G; Y, Y \in G' | Z, Z \in G'']
\end{aligned}$$

Unique measure on $\mathcal{F}$ consistent with information measures.

Information Measures ...

The Shannon Information Measure ...

I-Diagram: Venn-like graphic for I-measure relations

Entropy ~ area

Conditioning ~ area removal

Mutual information ~ intersection

Signed measure: Areas can represent $\mu^* < 0$.

However, area is zero, if $\mu^*(A) = 0, \; A \subset \Omega$.

Information Measures ...

The Shannon Information Measure ...

I-Diagram for two random variables:

$$X \sim \Pr(x)$$
$$Y \sim \Pr(y)$$
$$(X, Y) \sim \Pr(x, y)$$

Seven information measures:

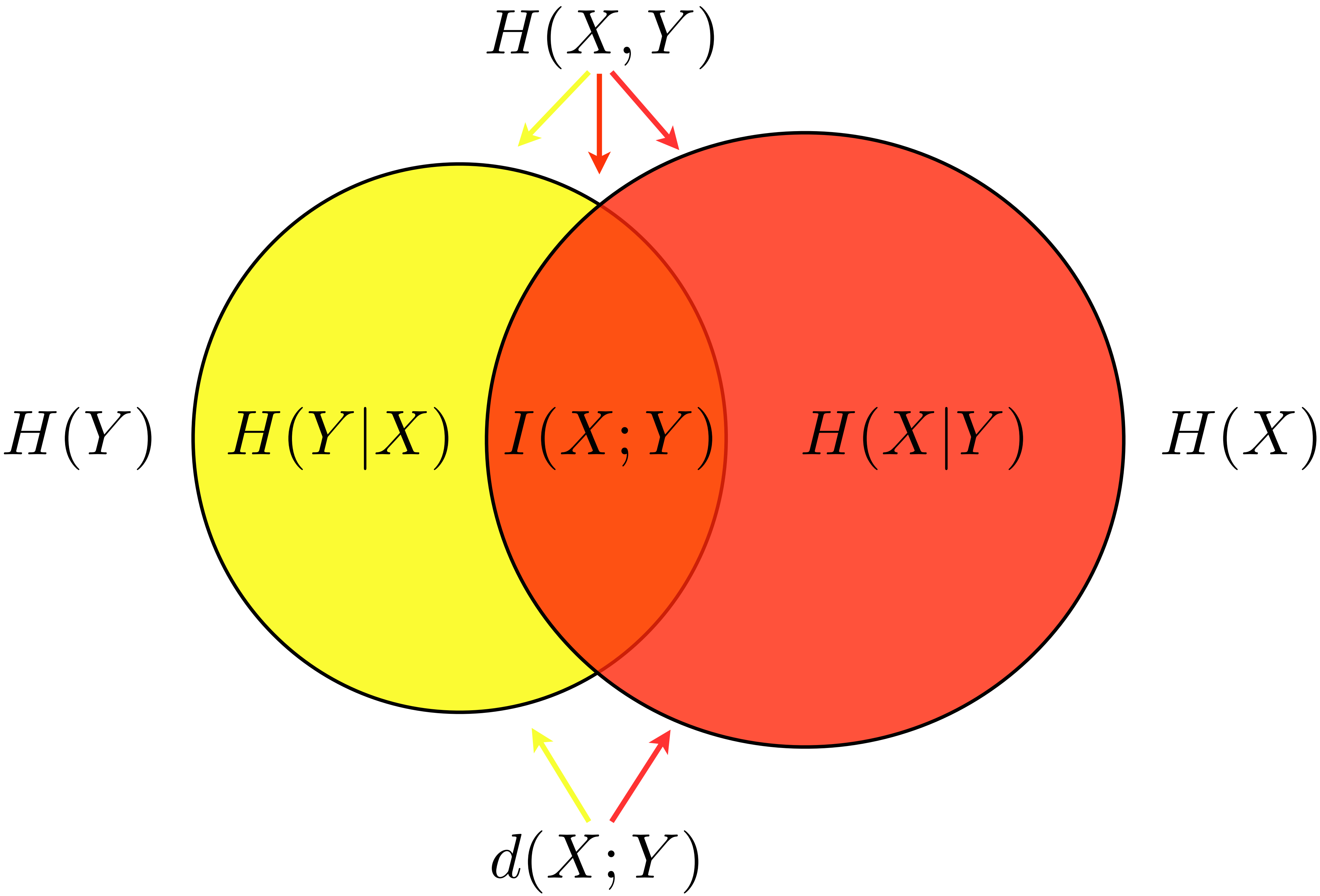
$$H[Y] \quad H[X] \quad H[X, Y]$$
$$H[X|Y] \quad I[X;Y] \quad H[Y|X] \qquad H[X|Y] + H[Y|X]$$

Three atoms:

$$H[X|Y] \quad I[X;Y] \quad H[Y|X]$$

Information Measures ...

Event Space Relationships of Information Quantifiers:



Information Measures ...

The I-Diagram ...

Three random variables:

$$X \sim \Pr(x)$$

$$Y \sim \Pr(y) \qquad (X, Y, Z) \sim \Pr(x, y, z)$$

$$Z \sim \Pr(z)$$

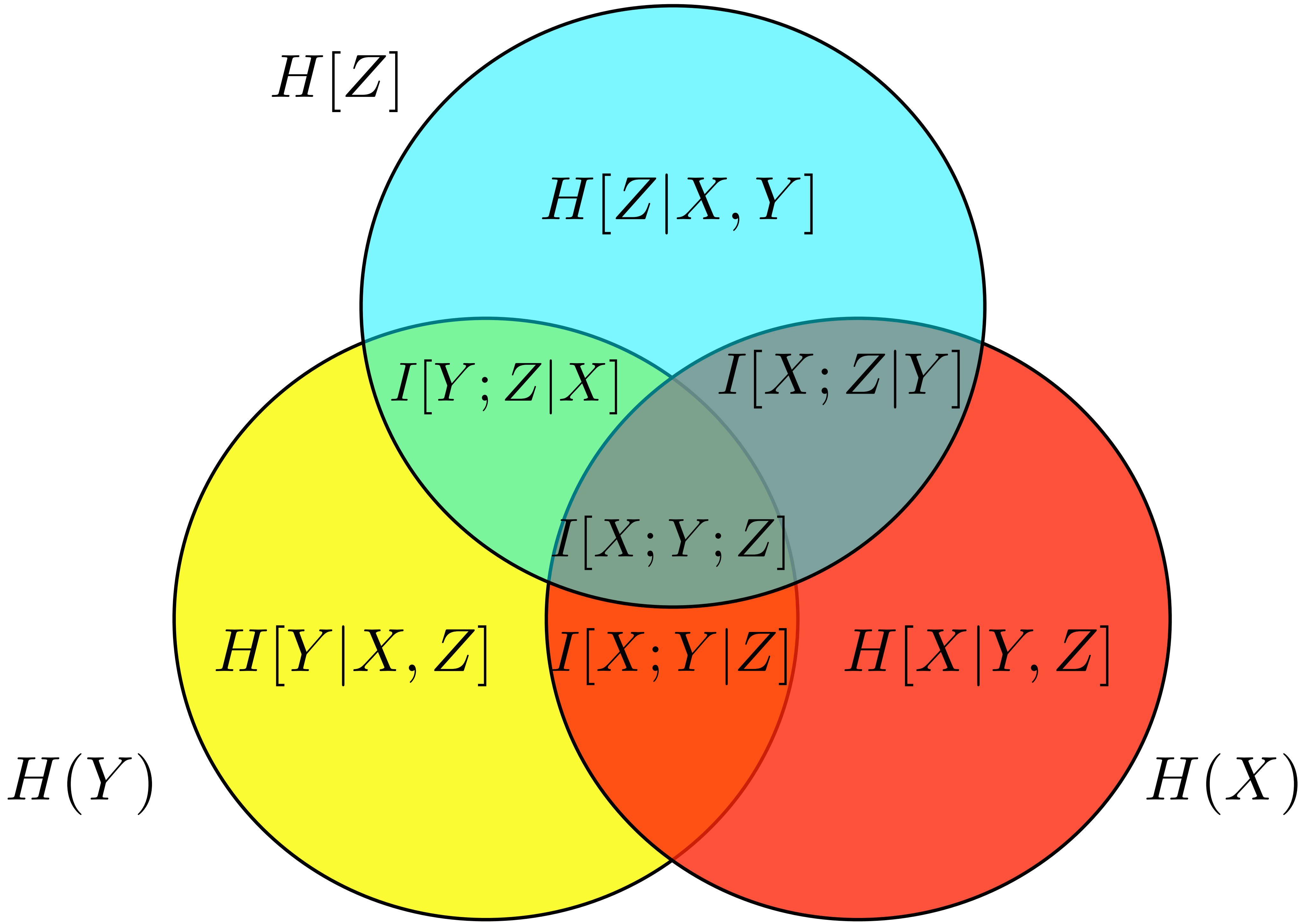
Information measures:

$$H[X] \quad H[Y] \quad H[Z] \quad \cdots \quad I[X;Y;Z] \quad \cdots \quad H[X,Y,Z]$$

7 atomic information measures.

Information Measures ...

Information diagram for three random variables:



Information Measures ...

Information diagram for three random variables ...

3-way mutual information is symmetric:

From diagram:

$$
\begin{aligned}
I[X;Y;Z] &= I[X;Y] - I[X;Y|Z] \\
&= I[Y;Z] - I[Y;Z|X] \\
&= I[X;Z] - I[X;Z|Y]
\end{aligned}
$$

Information Measures ...

Information diagram for three random variables ...

Three-way mutual information can be negative!

Consider:

$$X \sim \Pr(X = 0) = \Pr(X = 1) = 1/2$$

$$Y \sim \Pr(Y = 0) = \Pr(Y = 1) = 1/2$$

$$X \perp Y$$

$$Z = (X + Y) \mod 2$$

Recall RRXOR Process:

$$\begin{aligned} X_{t+2} \quad &= \quad (X_{t+1} + X_t) \mod 2, \\ X_t \quad \text{and} \quad & X_{t+1} \ \text{Bernoulli}(1/2) \end{aligned}$$

Information Measures ...

Information diagram for three random variables ...

Three-way mutual information can be negative!

Calculate:

$$I[X;Y] = 0$$

$$\begin{aligned} I[X;Y|Z] &= H[X|Z] - H[X|Y,Z] \\ &= 1 - 0 \\ &= 1 \end{aligned}$$

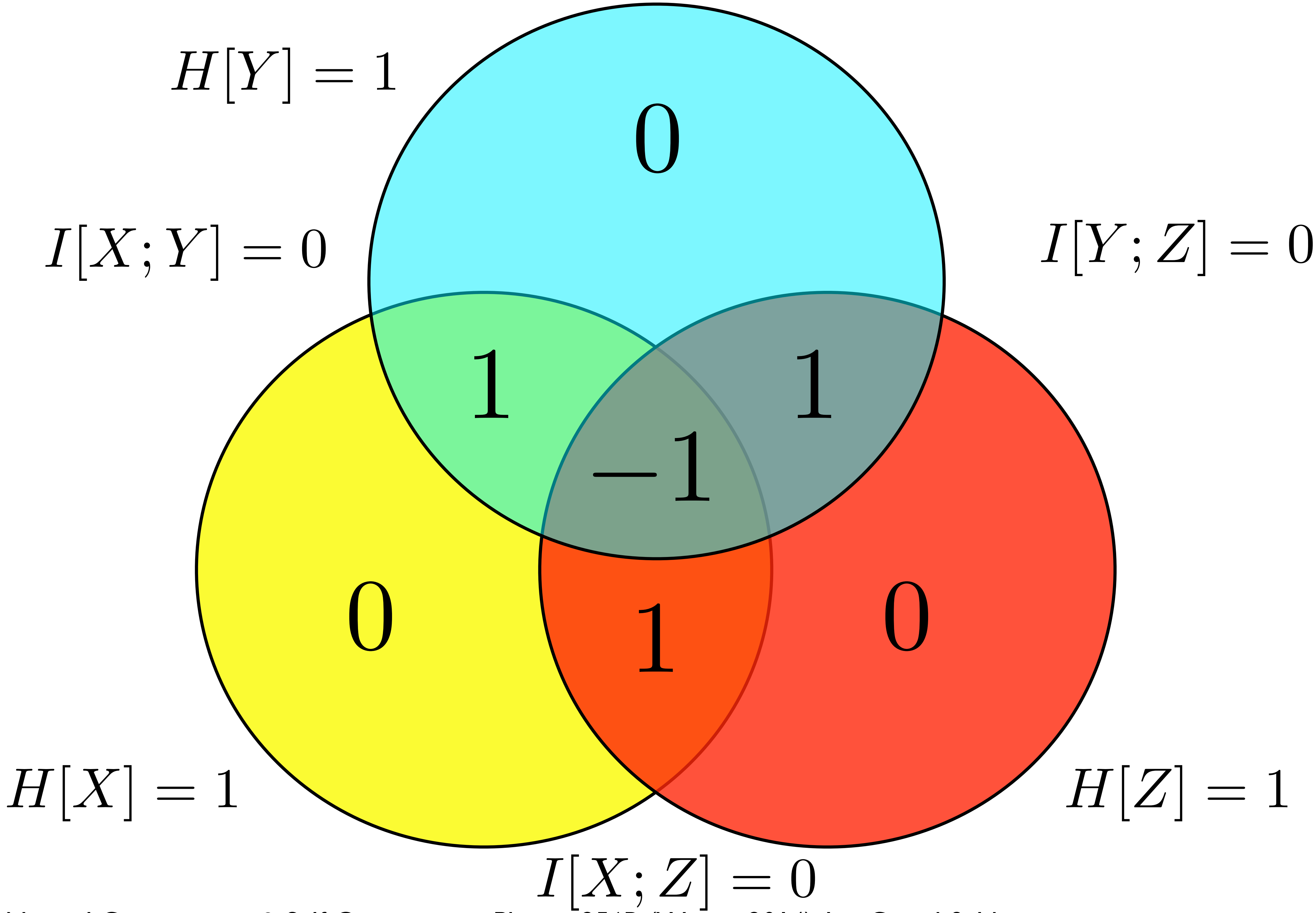
$$\begin{aligned} I[X;Y;Z] &= I[X;Y] - I[X;Y|Z] \\ &= -1 \end{aligned}$$

Shannon information measure is a signed measure!

Information Measures ...

Information diagram for three random variables ...

Three-way mutual information can be negative!

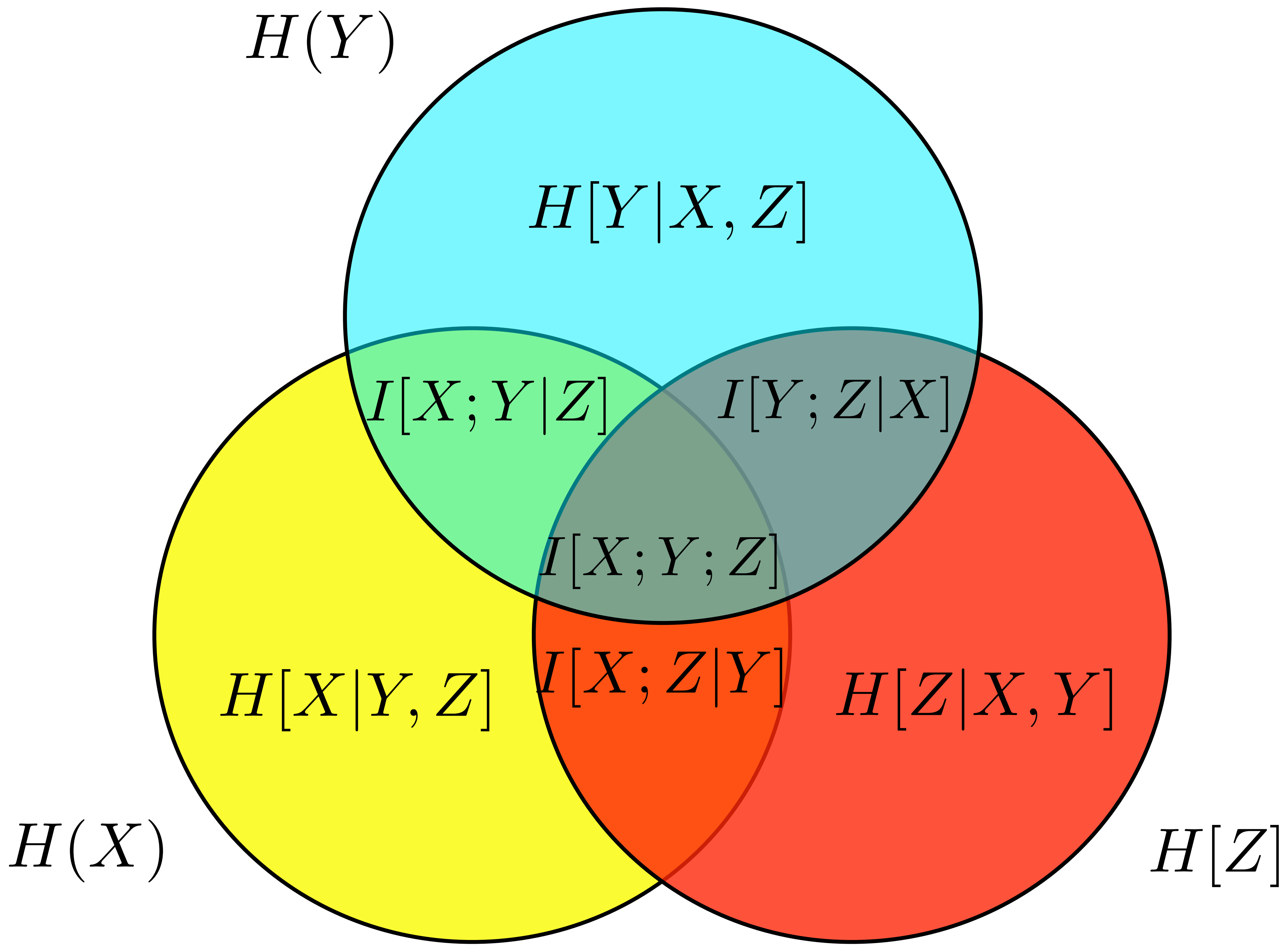


Information Measures ...

Information diagram for three random variables ...

Markov chain: $X \to Y \to Z$

Consequence of shielding: $I[X;Z|Y] = 0$



Information Measures ...

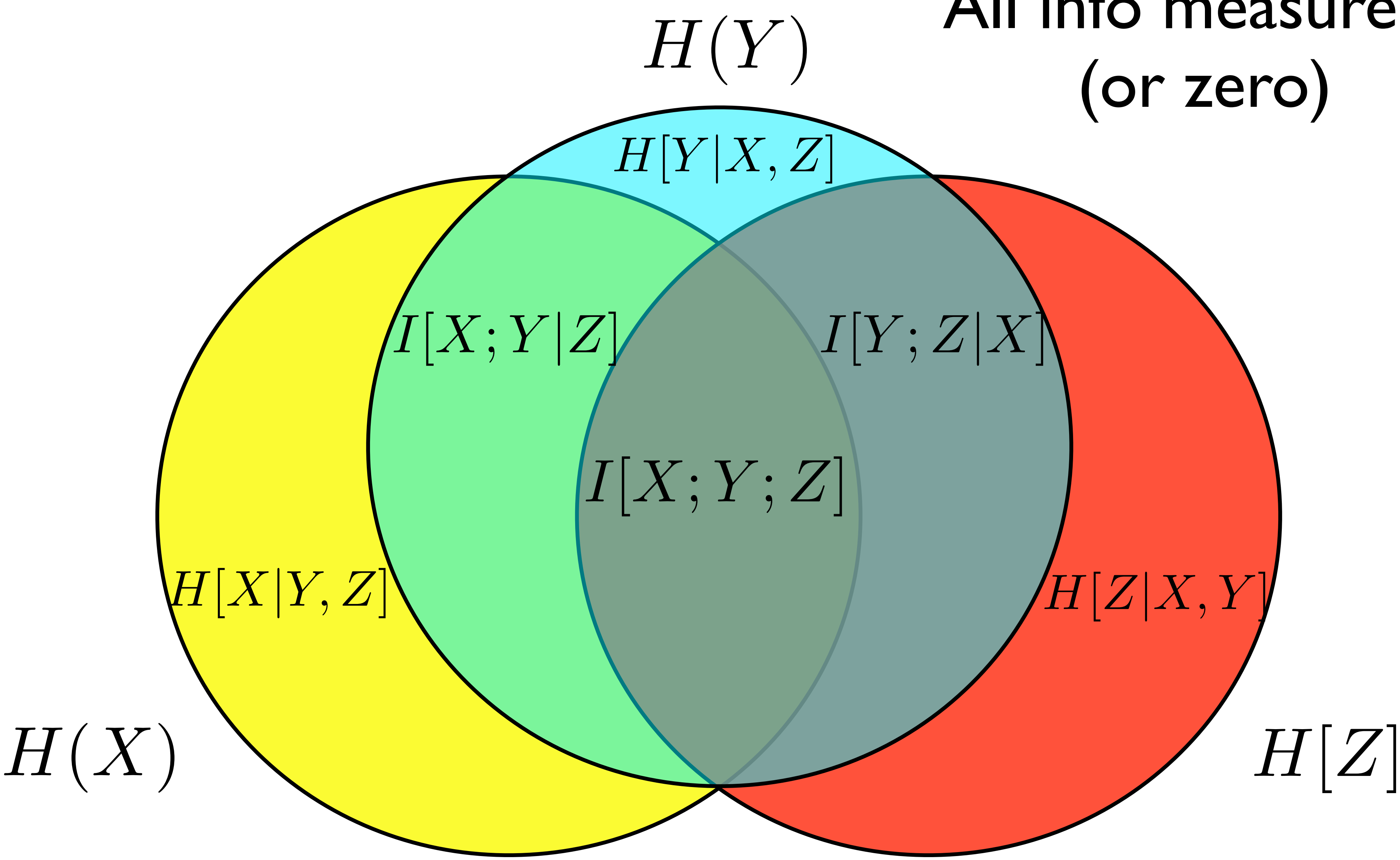
Information diagram for three random variables ...

Markov chain: $X \to Y \to Z$

Consequence of shielding: $I[X;Z|Y] = 0$

All areas positive

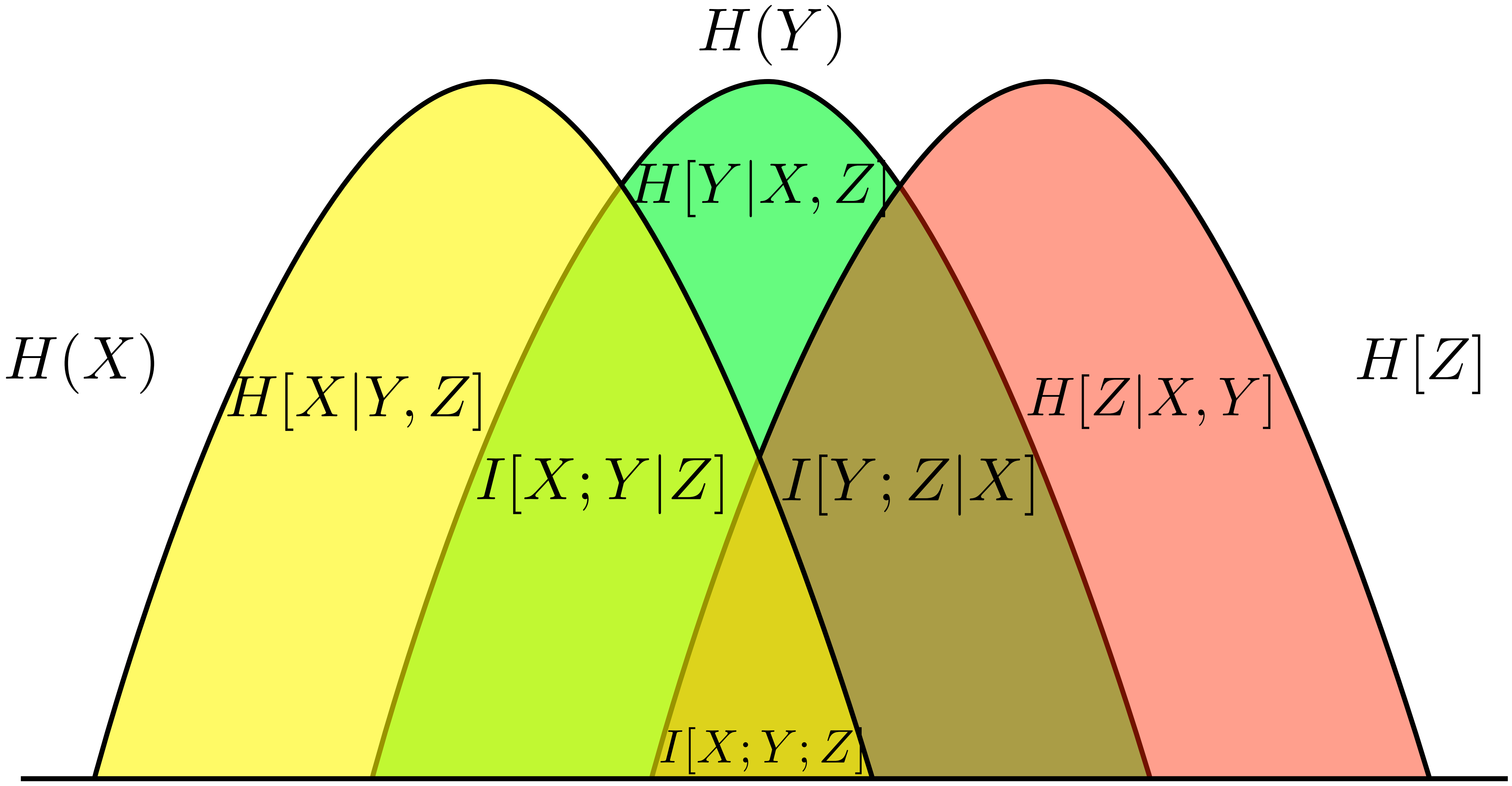
All info measures positive (or zero)



Information Measures ...

Information diagram for three random variables ...

Markov chain: $X \rightarrow Y \rightarrow Z$



Information Measures ...

Process I-diagrams:

Process has an infinite number of RVs!

$$\Pr(\overleftrightarrow{X}) = \Pr(\ldots, X_{-2}, X_{-1}, X_0, X_1, X_2, \ldots)$$

Rather:

$$\Pr\left(\overleftrightarrow{X}\right) = \Pr\left(\overleftarrow{X}\,\overrightarrow{X}\right)$$

Past as composite random variable: $\overleftarrow{X}$

Future as composite random variable: $\overrightarrow{X}$

Information Measures ...

Process I-diagrams ...

Start with 2-variable I-diagram and whittle down:

Information measures:

$$H[\overleftarrow{X}] \quad H[\overrightarrow{X}] \quad H[\overrightarrow{X}, \overleftarrow{X}]$$

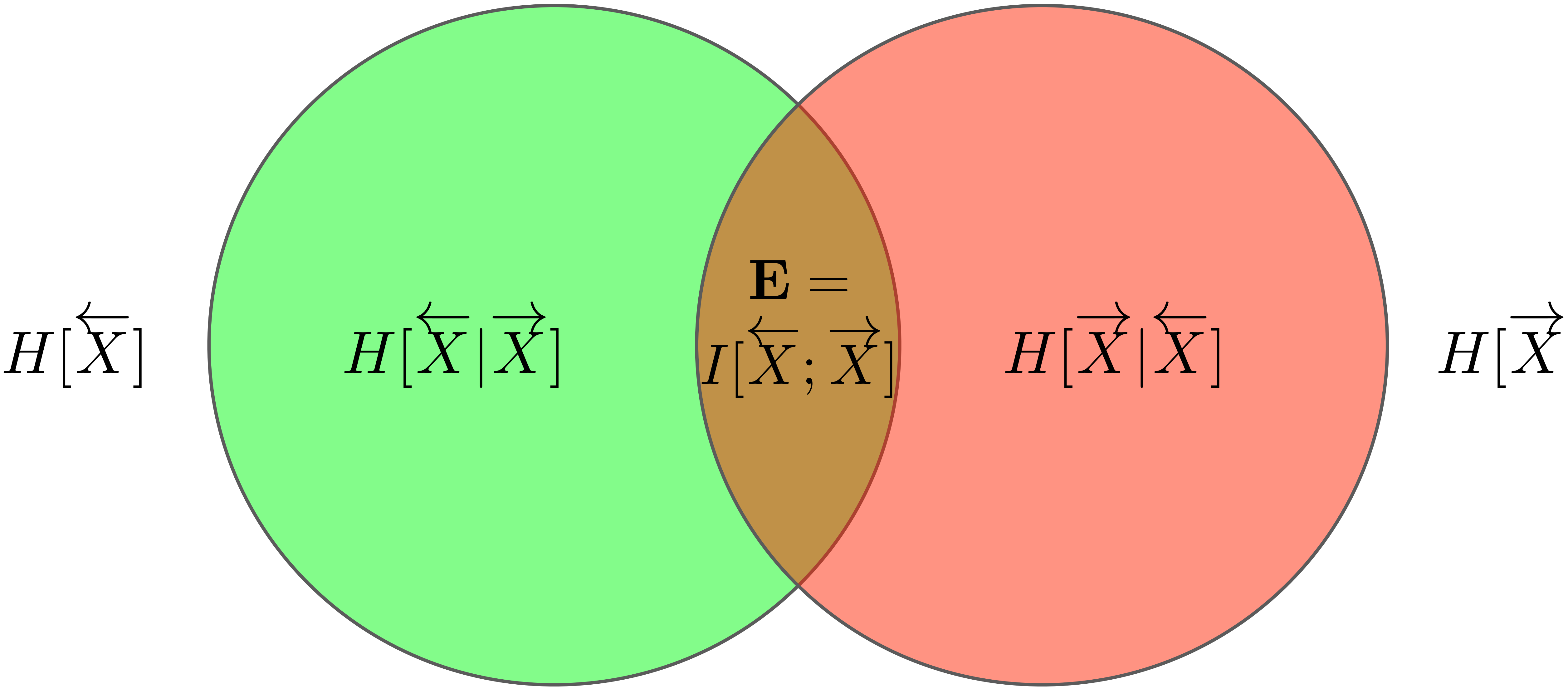
$$H[\overleftarrow{X}|\overrightarrow{X}] \quad H[\overrightarrow{X}|\overleftarrow{X}] \quad I[\overrightarrow{X}; \overleftarrow{X}] \quad H[\overrightarrow{X}|\overleftarrow{X}] + H[\overleftarrow{X}|\overrightarrow{X}]$$

There are $3 = 2^2\text{-}1$ atomic information measures:

$$H[\overrightarrow{X}|\overleftarrow{X}] \quad H[\overleftarrow{X}|\overrightarrow{X}] \quad I[\overrightarrow{X}; \overleftarrow{X}]$$

Information Measures ...

Process I-diagrams ...



Information Measures ...

Process I-diagrams ...

What is $H[\overleftarrow{X}|\overrightarrow{X}] + H[\overrightarrow{X}|\overleftarrow{X}]$?

Recall distance: $d(X,Y) = H[X|Y] + H[Y|X]$

So, $H[\overrightarrow{X}|\overleftarrow{X}] + H[\overleftarrow{X}|\overrightarrow{X}]$
is the distance between the past and the future!

Information Measures ...

Process I-diagrams ...

Nice picture (intuitive), but caution!

$H[\overleftarrow{X}]$ and $H[\overrightarrow{X}]$ are infinite for positive entropy-rate processes.

Rather, work with finite-L quantities, e.g,:

$$H[\overleftarrow{X}^L] \quad H[\overrightarrow{X}^L]$$

Then take limit: $\lim_{L \to \infty}$

Process I-diagrams ...

However, we do know that:

$$H[\overrightarrow{X}|\overleftarrow{X}] = \lim_{L \to \infty} H[\overrightarrow{X}^L|\overleftarrow{X}]$$

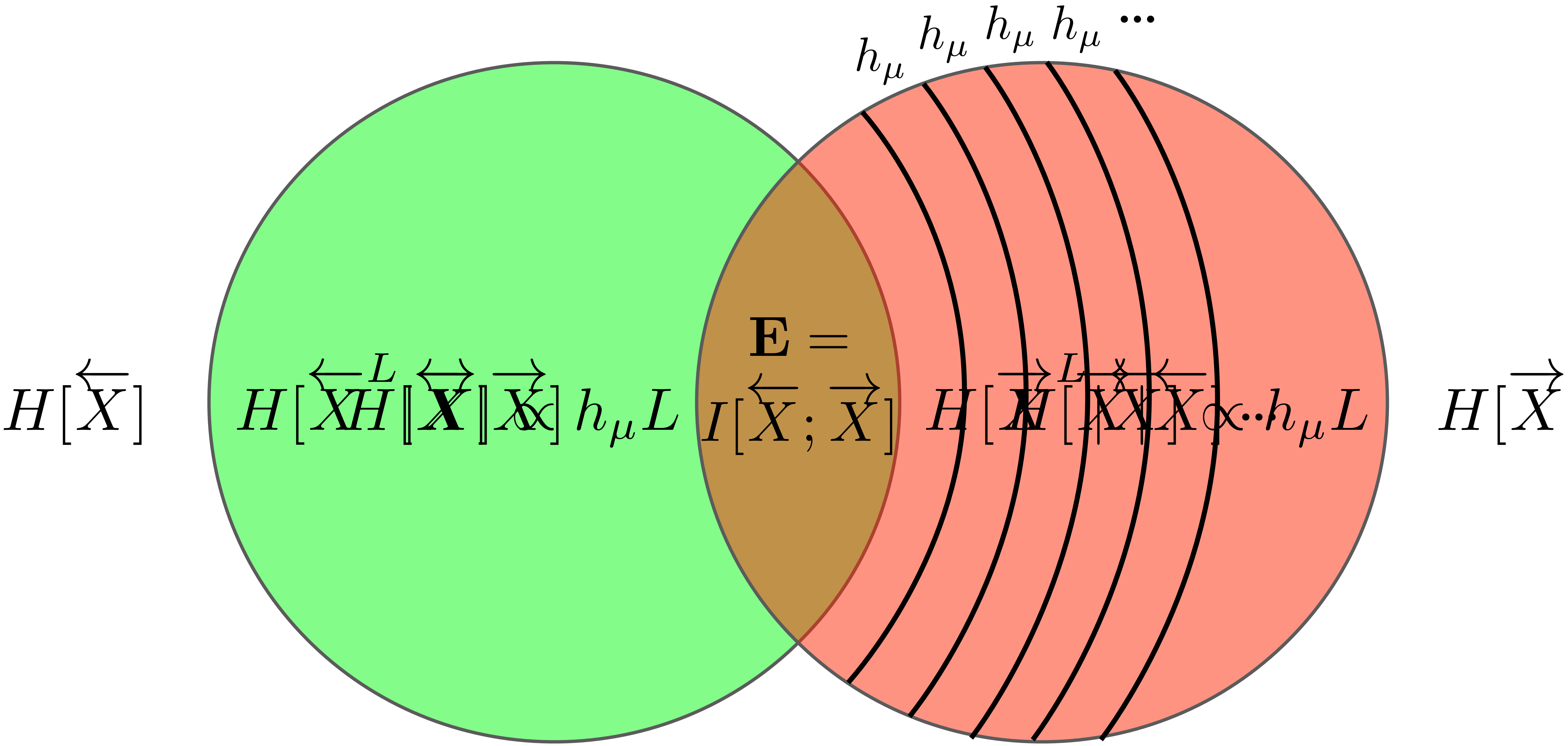
More to the point:

$$H[\overrightarrow{X}^L|\overleftarrow{X}] \propto h_\mu L$$

More accurate foliated I-diagram ...

Information Measures ...

Process I-diagrams ...



Information Measures ...

The Shannon Information Measure ...

For N > 3 random variables:

Higher dimensions with hyperspheres or

Unsymmetric and/or disconnected blobs in the plane.

The Shannon Information Measure ...

For N random variables:

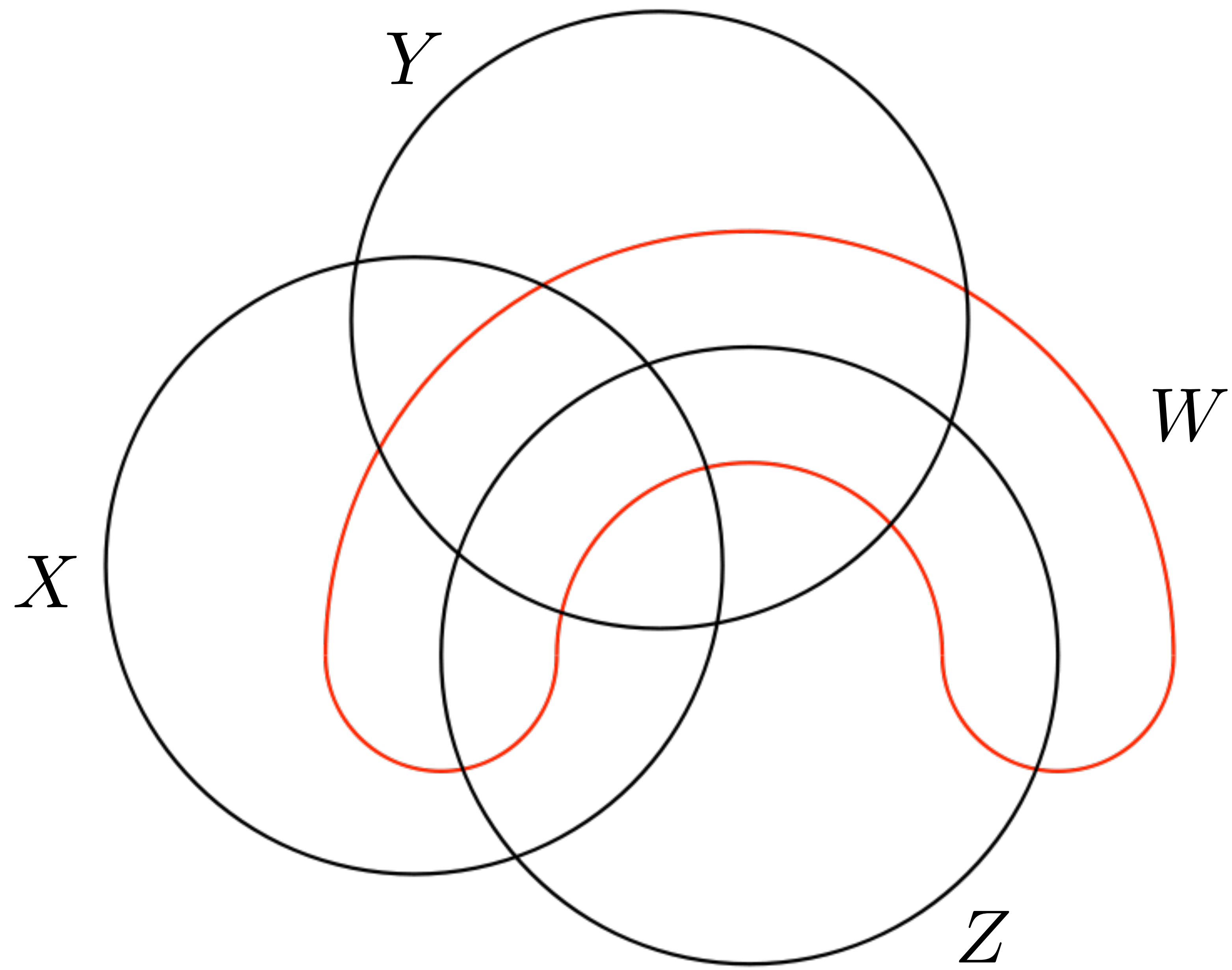
$\sim 2^N$ independent areas

$N > 3$ cannot be done with circles in the plane

Information Measures ...

The Shannon Information Measure ...

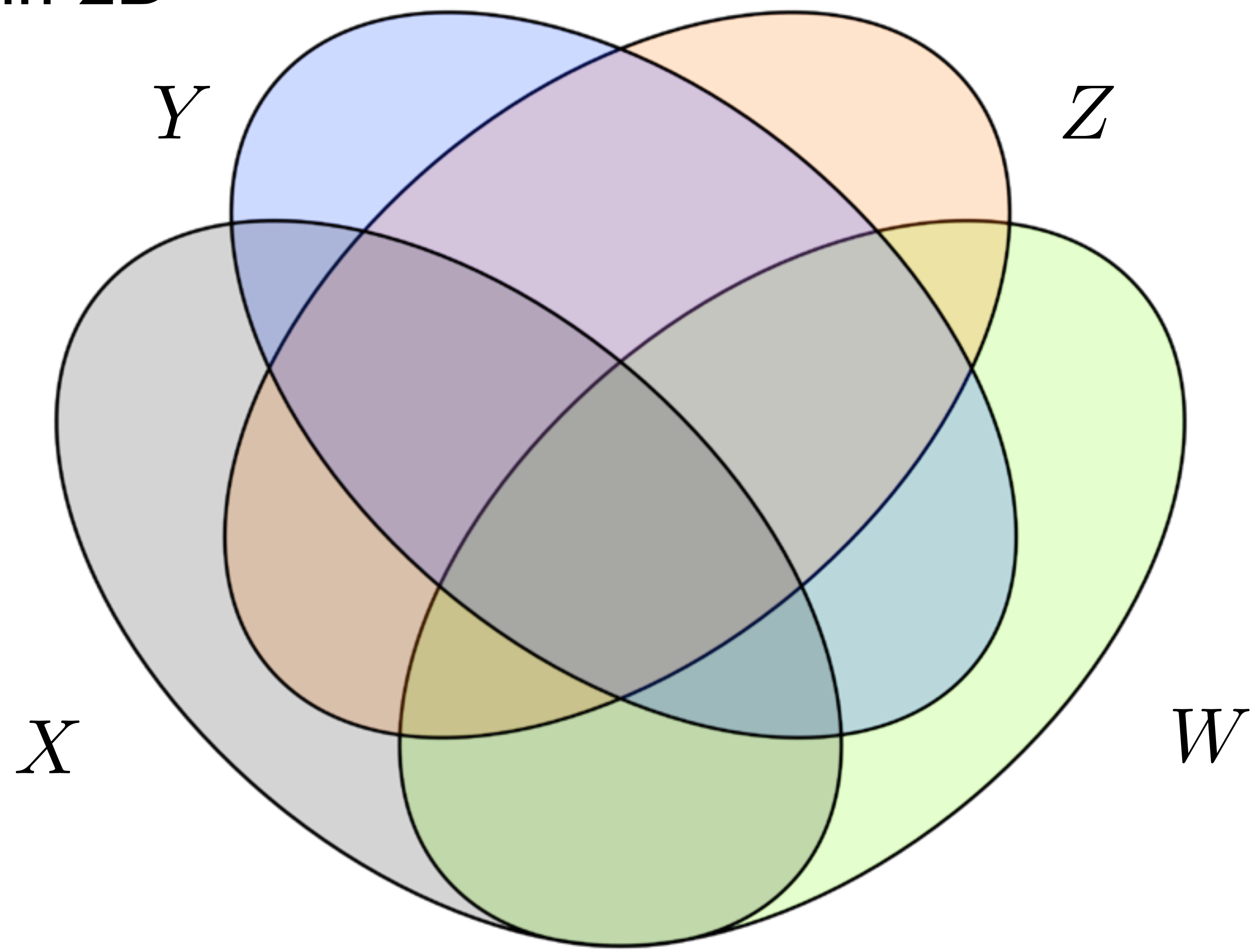
For N = 4 random variables: $(X, Y, Z, W) \sim \Pr(x, y, z, w)$
Three circles + sausage in 2D



Information Measures ...

The Shannon Information Measure ...

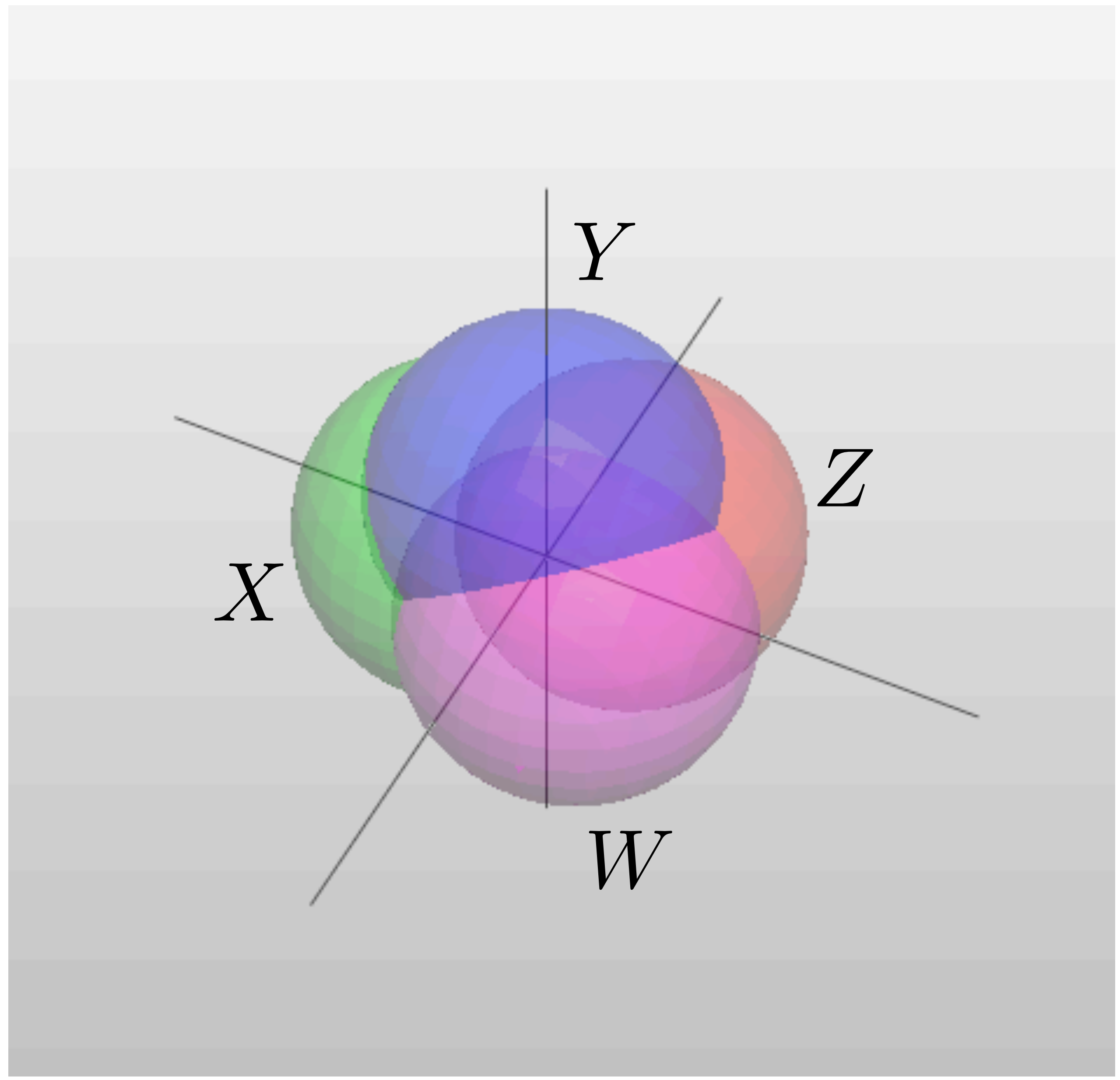
For N = 4 random variables:
Ellipses in 2D



Information Measures ...

The Shannon Information Measure ...

For N = 4 random variables:
Spheres in 3D



Information Measures ...

Multivariate Mutual Information(s):

$$(X_1, X_2, \ldots, X_N) \sim \Pr(x_1, x_2, \ldots, x_N)$$

What is N-way mutual information?

At least three different generalizations of 2-way.

1. Shared information to which all variables contribute:

$$I[X;Y] = H[X] + H[Y] - H[X,Y]$$

2. MI as relative entropy between joint and product of its marginals:

$$I[X;Y] = \mathcal{D}\left(\Pr(x,y) || \Pr(x)\Pr(y)\right)$$

3. Joint entropy minus all (single-variable) unshared information:

$$I[X;Y] = H[X,Y] - H[X|Y] - H[Y|X]$$

Information Measures ...

Multivariate Mutual Information(s) ...

Notation:

$$X_{0:N} = \{X_1, X_2, \ldots, X_{N-1}\}$$

Universal set over the variable indices: $\Omega_N = \{0, 1, \ldots, N-1\}$

Power set: $P(N) = \mathcal{P}\big(\Omega_N\big)$

Complement of $A \in P(N)$:

$$\bar{A} = \Omega_N \setminus A$$

Index set:

$$i \in A \subseteq \{1, \ldots, N\}$$

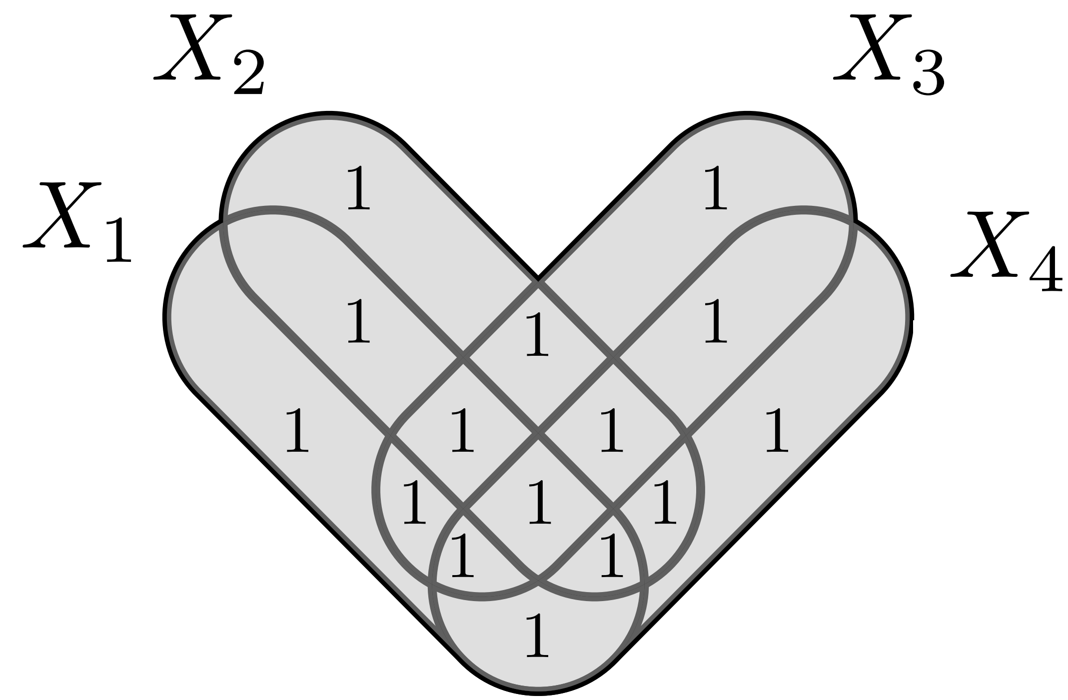
$$X_A = \{X_i : i \in A\}$$

Information Measures ...

Multivariate Mutual Information(s) ...

Joint entropy:

$$H[X_{0:N}] = - \sum_{\{x_{0:N}\}} \Pr(x_{0:N}) \log_2 \Pr(x_{0:N})$$



Gray level = # times atom counted

Multivariate Mutual Information(s) ...

Multivariate mutual information (aka Co-Information):

$$I[X_{0:N}] = I[X_1; X_2; \ldots; X_N]$$
$$= - \sum_{A \in P(N)} (-1)^{|A|} H[X_A]$$

Add and subtract all subset entropies:

$$I[X_0; X_1; X_2] = H[X_0] + H[X_1] + H[X_2]$$
$$- H[X_0, X_1] - H[X_0, X_2] - H[X_1, X_2]$$
$$+ H[X_0, X_1, X_2]$$

Information Measures ...

Multivariate Mutual Information(s) ...

Multivariate mutual information ...

$$I[X_{0:N}] = H[X_{0:N}] - \sum_{\substack{A \in P(N) \\ 0<|A|<N}} I[X_A | X_{\bar{A}}]$$

where, e.g.:

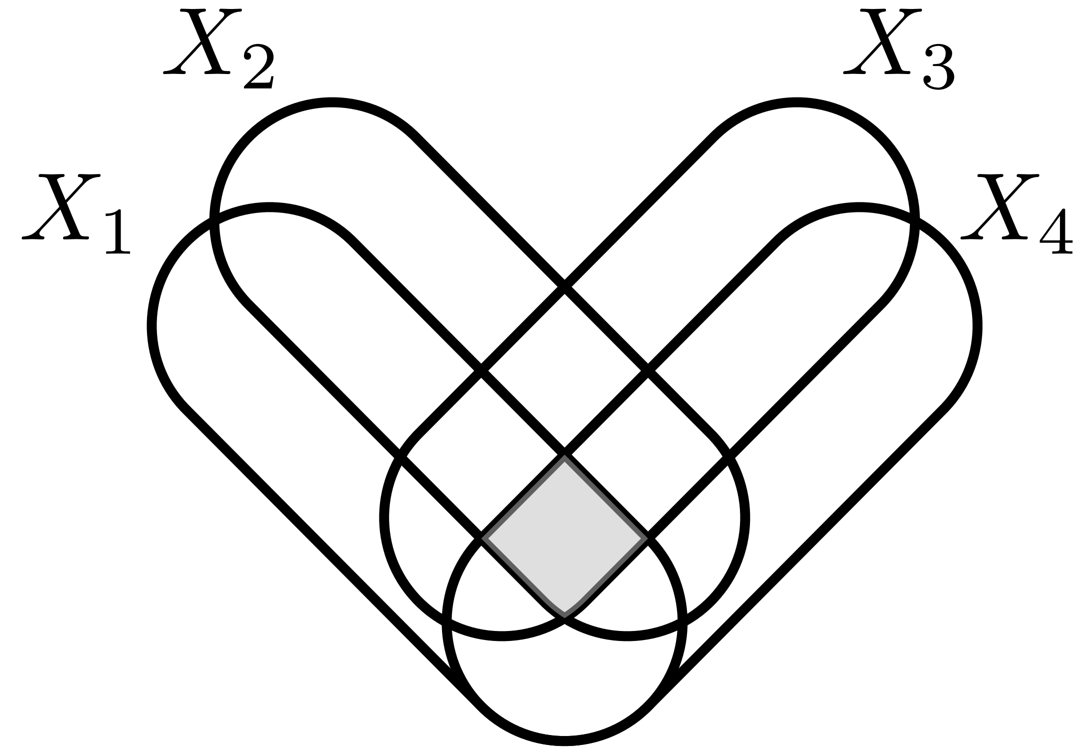
$$I[X_{\{1,3,4\}} | X_{\{0,2\}}] = I[X_1; X_3; X_4 | X_0, X_2]$$

$$I[X_{\{1\}} | X_{\{0,2\}}] = H[X_1 | X_0, X_2]$$

Information Measures ...

Multivariate Mutual Information(s) ...

Multivariate mutual information ...



Properties:

1. Common information to which all variables contribute.

2. Can be negative.

3. $I[X_{0:N}] = 0$, if any two variables are *completely* independent

(pairwise independent conditioned on any subset)

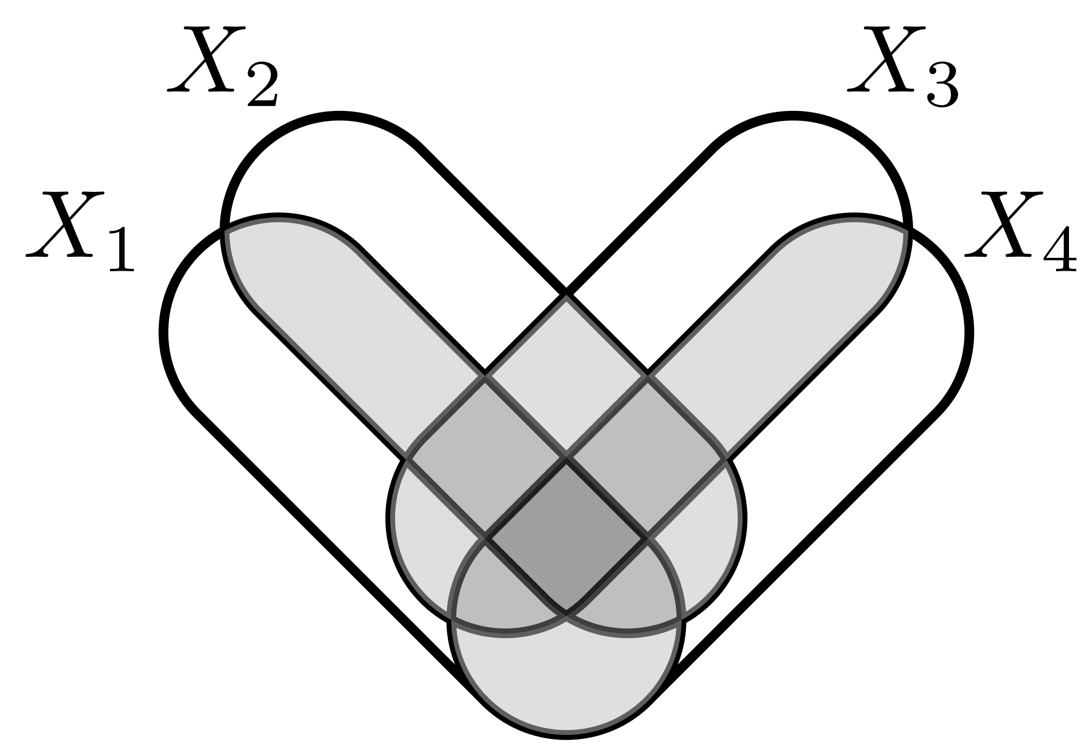
Information Measures ...

Multivariate Mutual Information(s) ...

Total correlation (aka Multi-information):

$$T[X_{0:N}] = \sum \Pr(x_{0:N}) \log_2 \left(\frac{\Pr(x_{0:N})}{\Pr(x_0) \ldots \Pr(x_N)} \right)$$

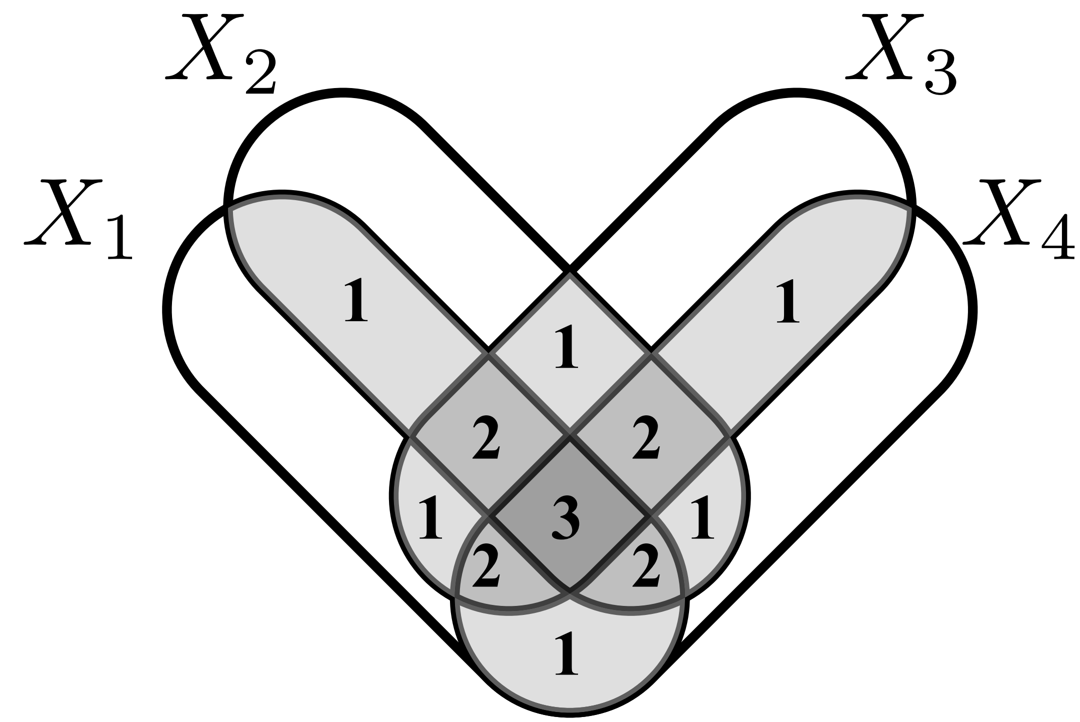
$$= \sum_{\substack{A \in P(N) \\ |A|=1}} H[X_A] - H[X_{0:N}]$$



Information Measures ...

Multivariate Mutual Information(s) ...

Total correlation ...



Properties:

1. $T[X_{0:N}] \geq 0$

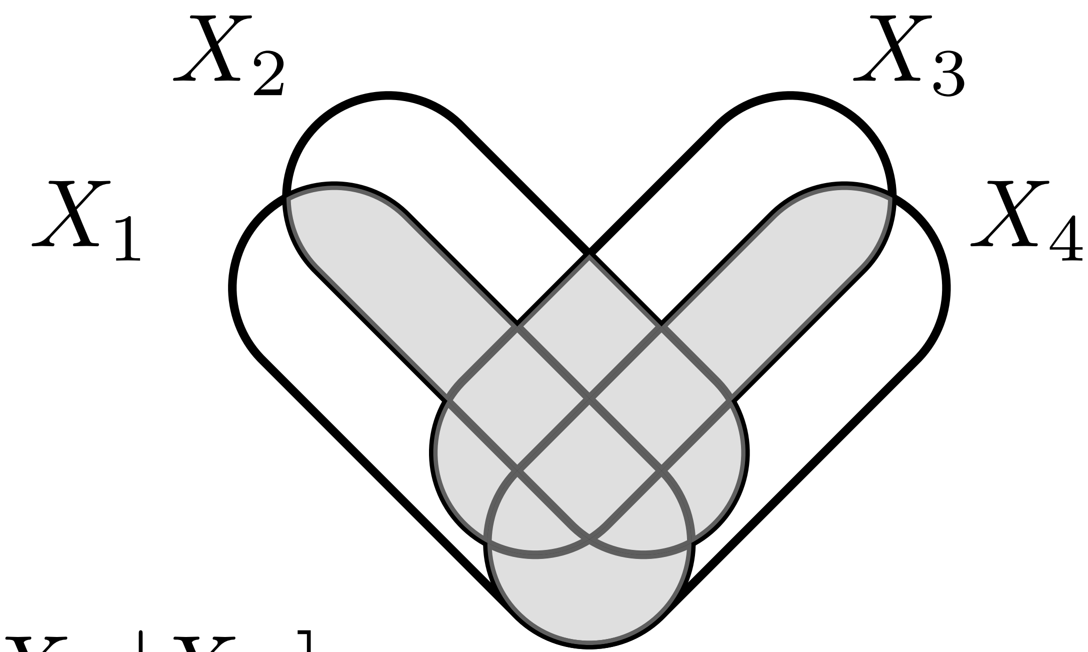
2. $X_0 \perp X_1, \ldots X_N \Rightarrow T[X_{0:N}] = T[X_{1:N}]$

3. Compares individuals to the entire set.

4. Includes redundant correlations.

Information Measures ...

Multivariate Mutual Information(s) ...



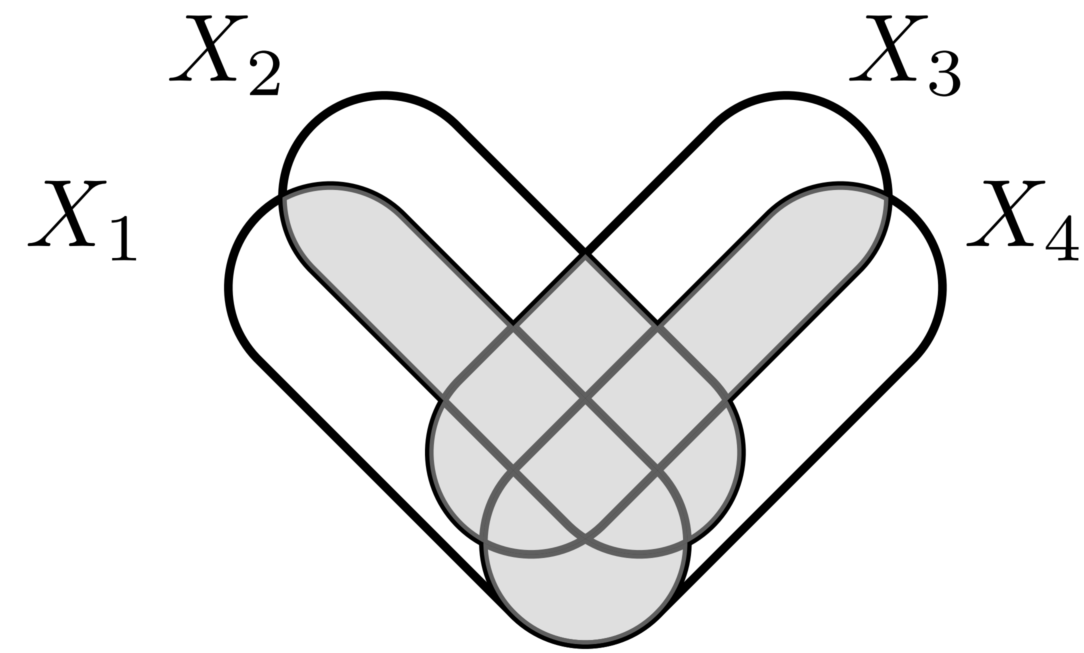
Bound Information:

$$B[X_{0:N}] = H[X_{0:N}] - \sum_{\substack{A \in P(N) \\ |A|=1}} H[X_A | X_{\bar{A}}]$$

MI as joint entropy minus all (single-variable) unshared information.

Information Measures ...

Multivariate Mutual Information(s) ...



Bound Information ...

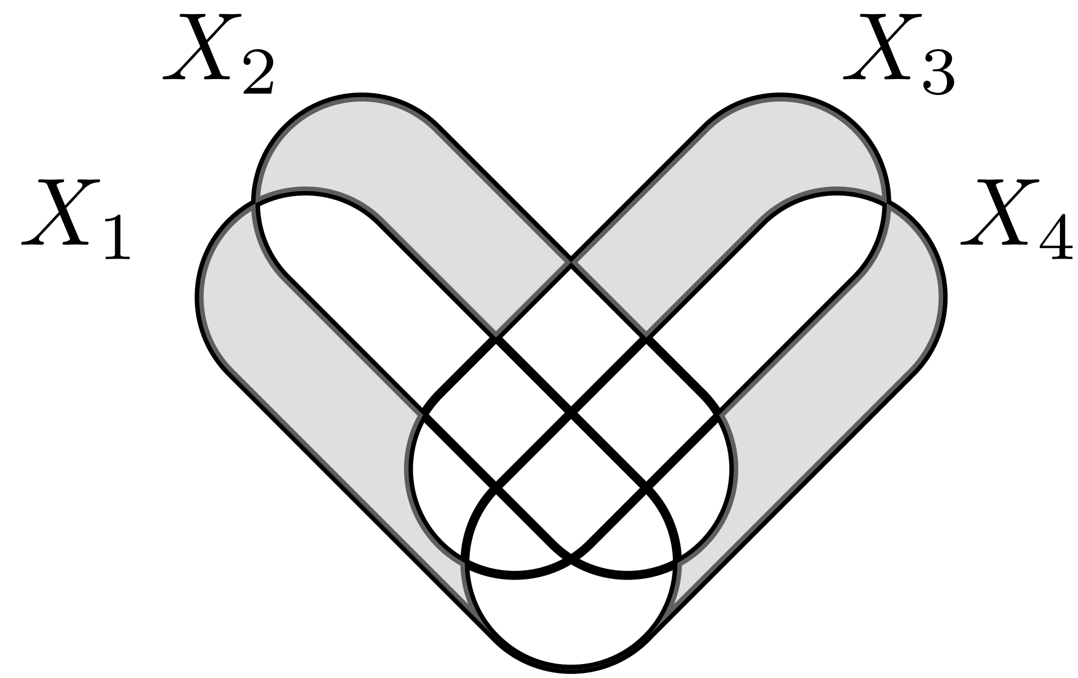
Properties:

1. $B[X_{0:N}] \geq 0$
2. $X_0 \perp X_1, \ldots X_N \Rightarrow B[X_{0:N}] = B[X_{1:N}]$

Multivariate Informations:

Residual Entropy (Anti-Mutual Information):

$$R[X_{0:N}] = H[X_{0:N}] - B[X_{0:N}]$$
$$= \sum_{\substack{A \in P(N) \\ |A|=1}} H[X_A | X_{\bar{A}}]$$



Information in individual variables that is not shared in any way.

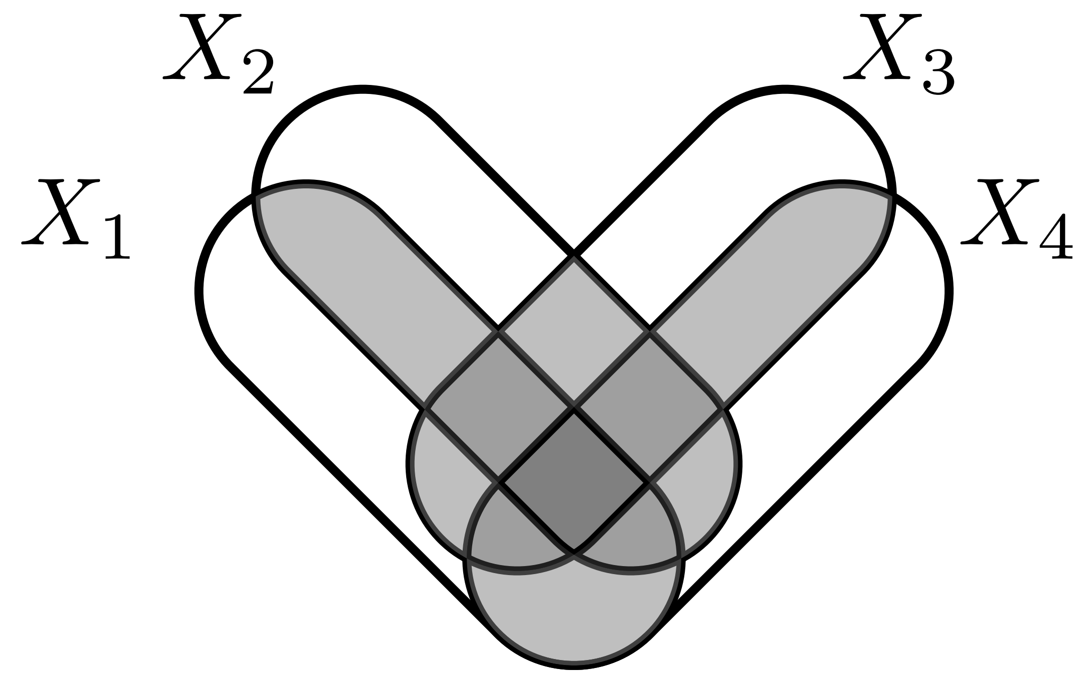
Total randomness localized to an individual variable and so not correlated to that in its peers.

Information Measures ...

Multivariate Informations:

Local Exogenous Information (Very Mutual Information):

$$W[X_{0:N}] = B[X_{0:N}] + T[X_{0:N}]$$
$$= \sum_{\substack{A \in P(N) \\ |A|=1}} I[X_A; X_{\bar{A}}]$$



Information in each variable that comes from its peers.

Discounts for randomness produced locally.

Information Measures ...

Reading for next lecture: CMR article *Anatomy*