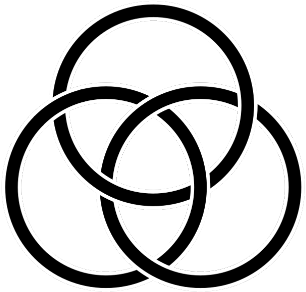
The information theory of higher-order interactions

From surprisal to Ising interactions, and beyond

Abel Jansma – DIEP, July 2023

University of Edinburgh
– Institute of Genetics and Cancer
– Higgs Centre for Theoretical Physics
– School of Informatics

@Abelaer



Outline

- **Goal: quantify higher-order structure.**
    - Information theory: *Multivariate entropy & mutual information*
    - Partial information decomposition: *Unique, redundant & synergistic information*
    - Statistical physics: *Interactions* in energy-based models
    - Are these related?
- Today:
    - Relating interactions in energy-based models to information theory.
    - Some ways in which synergy is better captured by these interactions than by entropy-based measures.

entropy MDPI

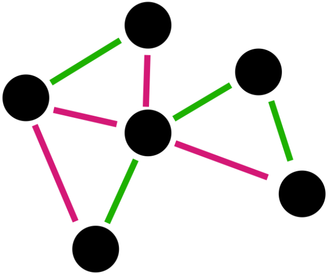
Article

Higher-Order Interactions and Their Duals Reveal Synergy and Logical Dependence beyond Shannon-Information

Abel Jansma [1,2,3]

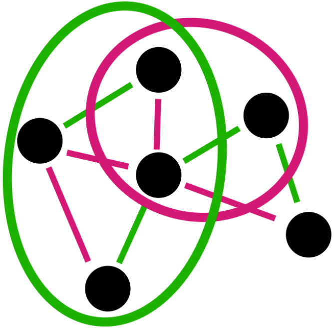
Higher-order structure

- Graph:



- Pairwise edges
- *e.g.* correlations, mutual information, regression, etc.

- Hypergraph:



- Higher-order edges
- Higher-order dependence?

Part 1: Higher-order information theory

- What is higher-order information?
- Why is it problematic?

Information and entropy

- Discrete random variable X, distributed as $p(X = x)$
- Question: how to quantify information/surprise S upon realisation $X = x$?
  - $S(X = x) = 0 \iff p(X = x) = 1$ (no surprise)
  - S decreases monotonically with $p(x)$
  - $X \perp\!\!\!\perp Y \implies S(X = x, Y = y) = S(X = x) + S(Y = y)$
- $\implies S(X = x) = -\log p(x)$
- Surprisal/Shannon information
- Expected surprise of X under $p(X)$: $H(X) = \mathbb{E}_p[S(x)] = -\sum_x p(x) \log p(x)$
- **Entropy** of X.

How to quantify dependence?

- Random variables X and Y
- Independence: $X \perp\!\!\!\perp Y \implies p(x, y) = p(x)p(y)$
- Question: *How far is the joint from the product of marginals?*
- Relative entropy/KL divergence: $D_{KL}\left(p(z) || q(z)\right) = \sum_z p(z) \log \frac{p(z)}{q(z)}$
- Mutual information:

$$
\begin{aligned}
MI(X; Y) &= D_{KL}(p(x, y) || p(x)p(y)) = \sum_{x,y} p(x, y) \log \frac{p(x, y)}{p(x)p(y)} \\
&= \sum_{x,y} p(x, y) \log p(x, y) - \sum_{x} p(x) \log p(x) - \sum_{y} p(y) \log p(y) \\
&= -H(X, Y) + H(X) + H(Y)
\end{aligned}
$$

- Conditional mutual information: $MI(X; Y|Z) = H(X|Z) + H(Y|Z) - H(X, Y|Z)$

Higher-order dependence

- *Difference between joint and product of marginals*

$$D_{KL}\left(p(x_1, x_2, \ldots, x_n) \;||\; \prod_{i=1}^{n} p(x_i)\right) = \sum_{i=1}^{n} H(X_i) - H(X_1, X_2, \ldots, X_n)$$
$$= \mathrm{TC}(X_1, X_2, \ldots, X_n)$$

- Problem: TC only considers terms of order 1 and n.
- Alternative: How does knowledge of X_3 affect $MI(X_1, X_2)$?

$$MI(X_1, X_2, X_3) = MI(X_1, X_2) - MI(X_1, X_2 \mid X_3)$$
$$= H(X_1) + H(X_2) + H(X_3) - H(X_1, X_2) - H(X_1, X_3) - H(X_2, X_3) + H(X_1, X_2, X_3)$$

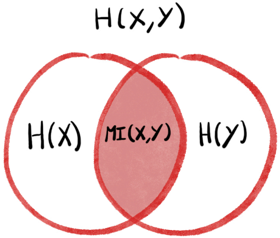
- (note: this is symmetric)
- Then continue inductively (up to minus sign):

$$MI(X_1, X_2, \ldots, X_n) = MI(X_1, X_2, \ldots, X_{n-1}) - MI(X_1, X_2, \ldots, X_{n-1} \mid X_n)$$

Information and set theory

- Mutual information as a Venn diagram:
- Question: How many elements in the intersection of finite sets A and B?

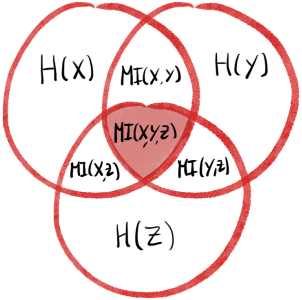
$$
\begin{aligned}
&| A \cup B | = | A | + | B | - | A \cap B | \\
&\implies | A \cap B | = | A | + | B | - | A \cup B | \\
&\implies MI(X, Y) = H(X) + H(Y) - H(X, Y)
\end{aligned}
$$



- Coincides with previous definition.
- Is information like a set measure?

Information and set theory

- Question: How many elements in the union of three finite sets A, B, and C?
- $|A \cup B \cup C| = |A| + |B| + |C| - |A \cap B| - |A \cap C| - |B \cap C| + |A \cap B \cap C|$
- Recall: $MI(X_1, X_2, X_3) = H(X_1) + H(X_2) + H(X_3) - H(X_1, X_2) - H(X_1, X_3) - H(X_2, X_3) + H(X_1, X_2, X_3)$
- Union instead of intersection?!
- Venn diagrams are misleading—entropy is not a measure!



- Intersection of two entropies is not an entropy.
- Mutual information can be negative.

Negative information

- XOR gate: $X_3 = X_1 \oplus X_2$

$$
\begin{aligned}
MI(X_1, X_2, X_3) &= MI(X_1, X_2) - MI(X_1, X_2 \mid X_3) \\
&= 0 - 1 = -1
\end{aligned}
$$

X_1	X_2	X_3
0	0	0
0	1	1
1	0	1
1	1	0

- **Problems:**
  - How to interpret negative information? (*Partial information decomposition*)
  - $MI(X_1, X_2, X_3)$ is bounded by pairwise quantities—can we separate the dependencies at different orders?
  - Why does the cardinality of a *union* of sets show up?
- To answer these questions, let's first look at a *different* approach to higher-order dependence.

Part 2: Higher-order interactions

- Two different perspectives on the Ising model
- A model-free solution to the inverse Ising problem

The Ising model: a physical perspective

- A model of interacting spins σ on a lattice, in a magnetic field h.
- $\sigma = \{\sigma_1, \ldots, \sigma_N\}, \quad \sigma_i \in \{0, 1\}$.
- The energy of a configuration—at equilibrium—is given by:

$$E(\sigma) = -\sum_{i,j} J_{ij}\sigma_i\sigma_j - \sum_i h_i\sigma_i$$

- High energy: $\uparrow \ \downarrow \ \uparrow \ \downarrow$
- Low energy: $\uparrow \ \uparrow \ \uparrow \ \uparrow$
- The probability of a configuration is given by:

$$p(\sigma) = \frac{1}{Z}\exp(-\beta E(\sigma))$$

- J_{ij} is called the *coupling*, or interaction, between spins i and j.
- Description of magnets, neurons, bird flocks, social dynamics, etc.

The Ising model: a statistical perspective (Jaynes ’57)

- Observe binary variables $\sigma = \{\sigma_1, \ldots, \sigma_N\}$.
- Write down a probability distribution $p(\sigma)$.
- Fewest assumptions: maximum entropy distribution

$$H(p) = -\sum_{\sigma} p(\sigma) \log p(\sigma)$$

- Subject to constraints $\sum_\sigma p(\sigma) = 1 \implies p(\sigma) = 2^{-N}$
- Add more constraints:

$$\sum_{\sigma} p(\sigma)\sigma_i = \mu_i, \quad \sum_{\sigma} p(\sigma)\sigma_i\sigma_j = \mu_{ij}$$

- $\implies p(\sigma) = \frac{1}{Z}\exp(-\sum_{i,j} J_{ij}\sigma_i\sigma_j - \sum_i h_i\sigma_i)$
- Ising model!
- Interactions and field fixed by observed moments.

Higher-order interactions & the inverse problem

- What if you constrain the higher-order moments?
- MaxEnt solution:

$$E(\sigma) = -\sum_i h_i \sigma_i - \sum_{i,j} J_{ij}\sigma_i\sigma_j - \sum_{i,j,k} J_{ijk}\sigma_i\sigma_j\sigma_k - \ldots$$

- An Ising model with **higher-order** interactions.
- Predicting properties of $p(\sigma)$ is the *forward* Ising problem.
- Fitting to data is the *inverse* Ising problem.
- Distentangles all orders of interactions.
- The inverse problem is hard.
  - MLE inference (exponential, pairwise only?)
  - Pseudolikelihood (polynomial, approximate but consistent, pairwise only?)
  - Restricted Boltzmann machines (approximate, unstable)

Interactions from the ground up

- What do we **really** mean when we say interaction? (Beentjes & Khamseh, 2020)
- *A difference which makes a difference.* (Bateson, 1972)
- A change in effect on an outcome, determined by the value of another variable, in the absence of other variables.

1-point interaction with respect to outcome Y:

$$I_i = \frac{\partial Y}{\partial X_i}\Big|_{\underline{X}=0} \qquad \underline{X} = X \setminus \{X_i\}$$

2-point interaction:

$$I_{ij} = \frac{\partial I_i}{\partial X_j}\Big|_{\underline{X}=0} = \frac{\partial^2 Y}{\partial X_j \partial X_i}\Big|_{\underline{X}=0} \qquad \underline{X} = X \setminus \{X_i, X_j\}$$

Interactions from the ground up

- A change in 2-point interaction is a *3-point interaction*:

$$I_{ijk} = \frac{\partial I_{ij}}{\partial X_k}\bigg|_{\underline{X}=0} = \frac{\partial^3 Y}{\partial X_k \partial X_j \partial X_i}\bigg|_{\underline{X}=0} \qquad \underline{X} = X \setminus \{X_i, X_j, X_k\}$$

- And so on.
- Interactions are defined with respect to an *outcome*.
- Now: let $Y = \log p(X)$, and $X \in \{0, 1\}^N$. Then:

$$\begin{aligned} I_i &= \frac{\partial \log p(X)}{\partial X_i}\bigg|_{\underline{X}=0} \\ &= \log p(X_i = 1 \mid \underline{X} = 0) - \log p(X_i = 0 \mid \underline{X} = 0) \\ &= \log \frac{p(X_i = 1 \mid \underline{X} = 0)}{p(X_i = 0 \mid \underline{X} = 0)} \end{aligned}$$

Interactions from the ground up

- Notation $p_{abc} = p(X_i = a, X_j = b, X_k = c \mid \underline{X} = 0)$
- 1-point interactions:

$$I_i = \frac{\partial \log p(X)}{\partial X_i}\Big|_{\underline{X}=0} = \log \frac{p_1}{p_0}$$

2-point:

$$I_{ij} = \frac{\partial^2 \log p(X)}{\partial X_j \partial X_i}\Big|_{\underline{X}=0} = \log \frac{p_{11}p_{00}}{p_{01}p_{10}}$$

3-point:

$$I_{ijk} = \frac{\partial^3 \log p(X)}{\partial X_k \partial X_j \partial X_i}\Big|_{\underline{X}=0} = \log \frac{p_{111}p_{100}p_{010}p_{001}}{p_{000}p_{011}p_{101}p_{110}}$$

- Symmetric in terms of the variables: $I_S = I_{\pi(S)}$
- Conditionally independent variables do not interact: $X_i \perp\!\!\!\perp X_j \mid \underline{X} \implies I_{ij} = 0$

Model-free interactions solve the inverse Ising problem!

$$
\begin{aligned}
E(X) &= -\sum_i h_i X_i - \sum_{i,j} J_{ij} X_i X_j - \sum_{i,j,k} J_{ijk} X_i X_j X_k - \ldots \\
I_{ijk} &= \left.\frac{\partial^3 \log p(X)}{\partial X_k \partial X_j \partial X_i}\right|_{\underline{X}=0} \\
&= -\left.\frac{\partial^3 E(X)}{\partial X_k \partial X_j \partial X_i}\right|_{\underline{X}=0} \\
&= J_{ijk} \\
&= \log \frac{p_{111} p_{100} p_{010} p_{001}}{p_{000} p_{011} p_{101} p_{110}} \approx \log \frac{\hat{n}_{111} \hat{n}_{100} \hat{n}_{010} \hat{n}_{001}}{\hat{n}_{000} \hat{n}_{011} \hat{n}_{101} \hat{n}_{110}}
\end{aligned}
$$

- $\hat{n}_{abc}$ is the number of samples that look like $(0, \ldots, 0, a, b, c, 0, \ldots, 0)$
- NB: This solves the *untruncated* problem.

Model-free interactions

- Surprisal of a state $X = x$: $-\log p(x)$
- Interactions are sums of surprisals:

$$
\begin{aligned}
I_i &= \log \frac{p_1}{p_0} = \log p_1 - \log p_0 \\
I_{ij} &= \log \frac{p_{11} p_{00}}{p_{01} p_{10}} = \log p_{11} + \log p_{11} - \log p_{01} - \log p_{10} \\
I_{ijk} &= \log \frac{p_{111} p_{100} p_{010} p_{001}}{p_{000} p_{011} p_{101} p_{110}} = \dots
\end{aligned}
$$

- What determines the alternating signs? (Even/odd)
- Similar to **mutual information**

Mutual information

- Higher-order mutual information:

$$MI(X, Y) = H(X) + H(Y) - H(X, Y)$$

$$MI(X, Y, Z) = H(X) + H(Y) + H(Z) - H(X, Y) - H(X, Z) - H(Y, Z) + H(X, Y, Z)$$

- Sign determined by even/odd number of variables
- *Inclusion/Exclusion principle*
- Higher-order structure is captured by **Möbius inversion**

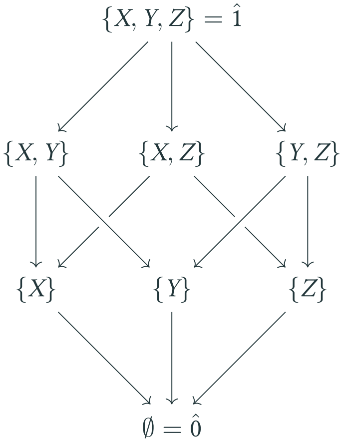
Part 3:
Higher-order structure as Möbius inversions

- The Möbius inversion formula
- Interactions and information as inversions
- Dualities

Möbius function

- Subsets form a lattice under inclusion:
- $S \leq T \iff S \subseteq T$
- Capture relationships in poset P: **Möbius function** $\mu_P : P \times P \to \mathbb{R}$

$$\mu_P(x, y) = \begin{cases} 1 & \text{if } x = y \\ -\sum\limits_{z : x \leq z < y} \mu_P(x, z) & \text{if } x < y \\ 0 & \text{otherwise} \end{cases}$$



Definition: Möbius inversion over a poset, Rota (1964)

Let P be a poset $(S, \leq)$, let $\mu_P : P \times P \to \mathbb{R}$ be the Möbius function, and let $g : P \to \mathbb{R}$ be a function on P. Then, the function

$$f(y) = \sum_{x \leq y} \mu_P(x, y) g(x)$$

is called the Möbius inversion of g on P. Furthermore, this can be inverted:

$$f(y) = \sum_{x \leq y} \mu_P(x, y) g(x) \iff g(y) = \sum_{x \leq y} f(x)$$

Möbius function: Examples

- Let $P = (\mathbb{N}, \leq)$, $\; f : P \to \mathbb{R}$, $\; F(x) = \sum_{y \leq x} f(y)$.

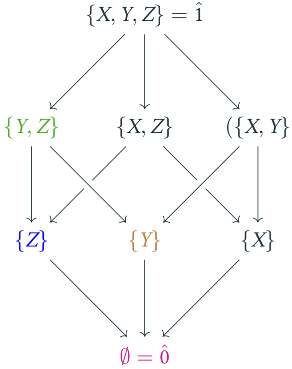
  $$\mu(a, a) = 1$$
  $$\mu(a, a+1) = -1$$
  $$\mu(a, a+k) = 0, \text{ for } k > 1$$

- Then the Möbius inversion of F is $f(x) = \sum_{y \leq x} \mu(y, x) F(y) = F(x) - F(x-1)$
- A discrete version of the fundamental theorem of calculus!
- If P is $\mathbb{N}$ ordered by divisibility, then μ is the inverse of the Riemann zeta function.
- Given an interval $[a, b]$ on a poset P, $\mu(a, b)$ is the *reduced Euler characteristic* of the associated simplicial complex.

Möbius functions on Boolean algebras

A powerset ordered by inclusion is a *Boolean algebra.*

$$
\begin{aligned}
\mu(\emptyset, \{Y, Z\}) &= - \sum_{\eta:\ \emptyset \le \eta < \{Y,Z\}} \mu(\emptyset, \eta) \\
&= -(\mu(\emptyset, \emptyset) + \mu(\emptyset, \{Y\}) + \mu(\emptyset, \{Z\})) \\
&= -(1 + \mu(\emptyset, \{Y\}) + \mu(\emptyset, \{Z\})) \\
&= -(1 - \mu(\emptyset, \emptyset) - \mu(\emptyset, \emptyset)) \\
&= -(1 - 1 - 1) = 1
\end{aligned}
$$



- On a Boolean algebra $\mu(x, y) = (-1)^{|y|-|x|}$
- The Möbius inversion of the intersection cardinality function $|*| : S \mapsto |\bigcap_i S_i|$ is the union cardinality.
- *e.g.* $|X \cup Y \cup Z| = |X| + |Y| + |Z| - |X \cap Y| - |X \cap Z| - |Y \cap Z| + |X \cap Y \cap Z|$
- **Exactly the sign-alternating sums we saw before!**

Möbius inversion

On a Boolean algebra P of variables T:

- Mutual information is the Möbius inversion of **marginal entropy**:

$$MI(\tau) = (-1)^{|\tau|-1} \sum_{\eta \leq \tau} \mu_P(\eta, \tau) H(\eta)$$

- Pointwise mutual information is the Möbius inversion of **marginal surprisal**:

$$\mathrm{pmi}(\tau) = (-1)^{|\tau|-1} \sum_{\eta \leq \tau} \mu_P(\eta, \tau) \log p(\eta)$$

- Model-free interactions are the Möbius inversion of **surprisal**:

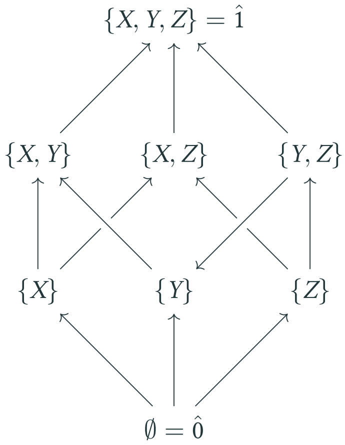
$$I(\tau; T) = \sum_{\eta \leq \tau} \mu(\eta, \tau) \log p(\eta = 1, T \setminus \eta = 0)$$

Dual quantities

- If $P = (S, \leq)$ is a lattice, then $P^{\mathrm{op}} = (S, \preceq)$ (where $a \preceq b \iff a \geq b$) is a lattice.
- What is dual mutual information $MI^*(\tau) = \sum_{\eta \preceq \tau} (-1)^{|\eta|+1} H(\eta)$?
- Dual MI of a single variable X:

$$\begin{aligned} MI^*(X) &= MI(X, Y, Z) - MI(Y, Z) \\ &= MI(Y, Z \mid X) = \Delta_X \end{aligned}$$

- *Conditional/differential* mutual information.
- $MI^*(X, Y) = H(X, Y, Z) - H(X, Y) = H(X \mid Y, Z)$
- In general context T: $MI^*(\tau) = MI(T \setminus \tau \mid \tau)$



Dual quantities

- Dual interactions $I^*(\tau; T) = \sum_{\eta \preceq \tau} (-1)^{|\eta|-|\tau|} \log p(\eta = 1, T \setminus \eta = 0)$
- Dual interaction of a single variable X in a system with 3 variables:

$$\begin{aligned} I^*(X; \{X, Y, Z\}) &= I(X, Y, Z) + I(Y, Z) \\ &= \log \frac{p_{111} p_{100}}{p_{110} p_{101}} \end{aligned}$$

- This is $I(Y, Z) \mid_{X=1}$.
- Dual interactions are interactions in a context of 1s:
- $I^*(\tau; T) = I(T \setminus \tau) \mid_{\tau=1}$
- **Outeractions**

Summary

- Mutual information is the Möbius inversion of marginal entropy.
- Pointwise mutual information is the Möbius inversion of marginal surprisal.
- Model-free interactions are the Möbius inversion of surprisal.
- Dual mutual information is conditional mutual information.
- Dual interactions are interactions in a context of 1s.
- **NB:** These all imply an intuitive inverse relation:

$$f(y) = \sum_{x \leq y} \mu_P(x, y) g(x) \iff g(y) = \sum_{x \leq y} f(x)$$

Summary

- Define: $eval_T : \log p(R = r) \mapsto \log p(R = 1, T \setminus R = 0)$
- Then:

$$
\begin{tikzcd}
MI^*(R) = MI(T \setminus R \mid R) & -H(R) \arrow[l, mapsto, "M_{pop}"'] \arrow[r, mapsto, "M_P"] & MI(R) \\
\mathrm{pmi}^*(R = r) \arrow[u, mapsto, "\mathbb{E}"'] \arrow[d, mapsto, "eval_T"] & \log p(R = r) \arrow[l, mapsto, "M_{pop}"'] \arrow[r, mapsto, "M_P"] \arrow[u, mapsto, "\mathbb{E}"'] \arrow[d, mapsto, "eval_T"] & \mathrm{pmi}(R = r) \arrow[u, mapsto, "\mathbb{E}"'] \arrow[d, mapsto, "eval_T"] \\
I^*(R; T) & \log p(R = 1; T = 0) \arrow[l, mapsto, "M_{pop}"'] \arrow[r, mapsto, "M_P"] & I(R; T)
\end{tikzcd}
$$

Example: $T = \{X, Y, Z\}$ and $R = \{X, Y\}$

$$\begin{tikzcd}
\begin{array}{c}\sum_{(x,y,z)\in\mathcal{X}\times\mathcal{Y}\times\mathcal{Z}} p(x,y,z)\log p(x,y,z) \\ -\sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} p(x,y)\log p(x,y)\end{array} & \sum_{(x,y)\in X\times Y} p(x,y)\log p(x,y) \arrow[l, mapsto, "M_{pop}"'] \arrow[r, mapsto, "M_P"] & \begin{array}{c}\sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} p(x,y)\log p(x,y) \\ -\sum_{x\in\mathcal{X}} p(x)\log p(x) \\ -\sum_{y\in\mathcal{Y}} p(y)\log p(y)\end{array} \\
\log\frac{p(x,y,z)}{p(x,y)} \arrow[u, mapsto, "\mathbb{E}"'] \arrow[d, mapsto, "eval_T"'] & \log p(x,y) \arrow[l, mapsto, "M_{pop}"'] \arrow[r, mapsto, "M_P"] \arrow[u, mapsto, "\mathbb{E}"'] \arrow[d, mapsto, "eval_T"'] & \log\frac{p(x,y)p(\emptyset)}{p(x)p(y)} \arrow[u, mapsto, "\mathbb{E}"'] \arrow[d, mapsto, "eval_T"'] \\
\log\frac{p(1,1,1)}{p(1,1,0)} & \log p(1,1,0) \arrow[l, mapsto, "M_{pop}"'] \arrow[r, mapsto, "M_P"] & \log\frac{p(1,1,0)p(0,0,0)}{p(1,0,0)p(0,1,0)}
\end{tikzcd}$$

Part 4: Results

- What do nonzero interactions correspond to?
- Logic gates, causal dynamics, and di/triadic distributions.
- How do interactions differ from information?

Results: Synergy in logic gates

- What does a 3-pt interaction correspond to?

$$I_{ABC} = \log \frac{p_{111}p_{100}p_{010}p_{001}}{p_{000}p_{011}p_{101}p_{110}}$$

- Maximally positive $\implies$ only terms in numerator are > 0.

A	B	C
0	0	1
0	1	0
1	0	0
1	1	1

- XNOR gate!
- (XOR is maximally negative)

Results: Synergy in logic gates

- XOR is *synergistic*: Knowing all but one of the input bits gives zero information on output.
- It is symmetric!
- Can be generalised to n-point XOR:

$$X_{n+1} = \sum_{i=1}^{n} X_i \mod 2$$

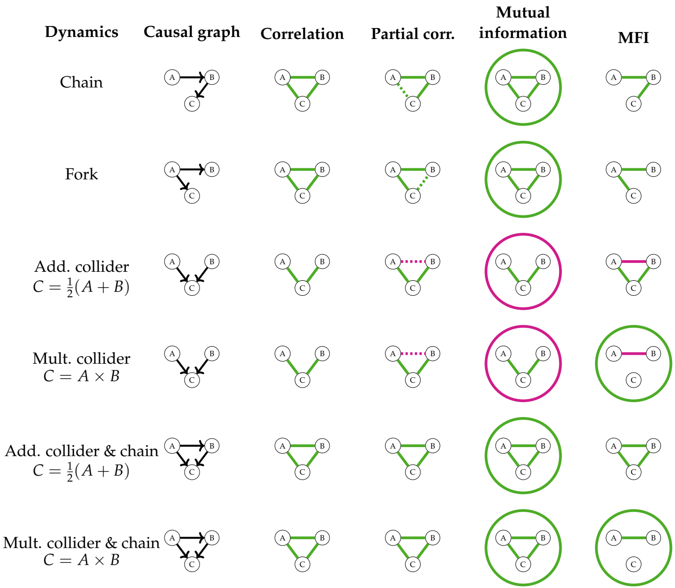
- Still maximally synergistic
- Still symmetric (consider $X_i \leftrightarrow X_{n+1}$)
- Still maximal 3-point interaction (parity determines sign)

Results: Synergy in logic gates

- Let $p(\text{allowed state}) = p$ and $p(\text{forbidden state}) = \epsilon$.
- Let $I = 4 \log \frac{p}{\epsilon}$
- Interactions have higher resolution than MI.
- AND$\sim$NOR and OR$\sim$NAND.
- Def. $\boldsymbol{J_A} = \boldsymbol{I_{ABC}} - \boldsymbol{I_{BC}}$
- $\boldsymbol{J_A}$ has perfect resolution.
- $\boldsymbol{J_A}^{\text{XNOR}} > \boldsymbol{J_A}^{\text{NOR}} > \boldsymbol{J_A}^{\text{AND}}$.
- Ordered by synergistic content.

$\mathcal{G}$	$\boldsymbol{I_{ABC}}$	$\boldsymbol{MI_{ABC}}$	$\boldsymbol{J_A}$
XNOR	I	-1	$\frac{3}{2}I$
XOR	$-I$	-1	$-\frac{3}{2}I$
AND	$\frac{1}{2}I$	-0.189	$\frac{1}{2}I$
OR	$-\frac{1}{2}I$	-0.189	$-I$
NAND	$-\frac{1}{2}I$	-0.189	$-\frac{1}{2}I$
NOR	$\frac{1}{2}I$	-0.189	I

Results: Causal dynamics



Results: Dy- and Triadic distribution

- 6 variables: $(X_0, X_1, Y_0, Y_1, Z_0, Z_1)$
- Dyadic: $X_0 = Y_1, Y_0 = Z_1, Z_0 = X_1$
- Triadic: $X_0 + Y_0 + Z_0 = 0 \mod 2$ and $X_1 = Y_1 = Z_1$
- Variables combined to form categorical variables **X, Y, Z**.
- $(X_0, X_1) = (1, 1) \implies \mathbf{X} = 3$
- Indistinguishable by almost all information measures. (James & Crutchfield, 2017)
- PID: has to identify in- and output variables.
- Symmetrised categorical interactions: $\mathbf{I}$
  - Dyadic: $\mathbf{I}_{XYZ} = \log 1 = 0$
  - Triadic: $\mathbf{I}_{XYZ} = 64 \log \frac{\epsilon}{p}$

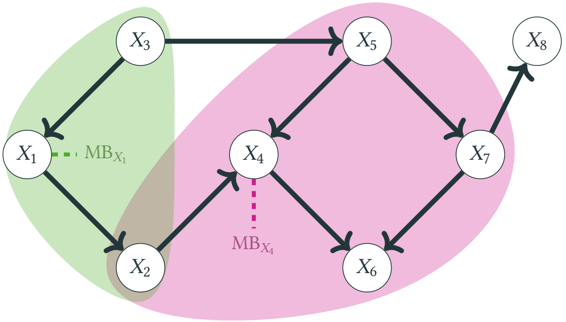
Dyadic				Triadic			
X	**Y**	**Z**	P(s)	**X**	**Y**	**Z**	P(s)
0	0	0	p	0	0	0	p
0	2	1	p	1	1	1	p
1	0	2	p	0	2	2	p
1	2	3	p	1	3	3	p
2	1	0	p	2	0	2	p
2	3	1	p	3	1	3	p
3	1	2	p	2	2	0	p
3	3	3	p	3	3	1	p
else			ϵ	else			ϵ

Part 5: Conclusion

- Teaser: model-free interactions on real data
- Conclusion

Teaser: model-free interactions on real data

- To estimate $p(X_i = \alpha, X_j = \beta \mid \underline{X} = 0)$, count samples: $X = (0, 0, \ldots, \alpha, \ldots, \beta, \ldots, 0)$
- Trick: find a set $\mathrm{MB}_{X_i} \subset \underline{X}$ s.t. $p(X_i \mid \underline{X}) = p(X_i \mid \mathrm{MB}_{X_i})$



- MB: Markov Blanket
- $X_i \perp\!\!\!\perp X \setminus \mathrm{MB}_{X_i} \mid \mathrm{MB}_{X_i}$
- d-separates X_i from ‘the rest’
- Finding minimal MB is ‘Causal Rung 2-hard’

Teaser: model-free interactions on real data

- Estimation on real data becomes tractable using Markov blankets.
- Error from MB estimate is bounded by pointwise mutual information (a good thing)
- MFIs in gene expression data: 1000 genes, 20k cells.
- Interactions at up to seventh order.
- $I_{G_1,G_2,G_3,\ldots} \neq 0 \implies$ non-random gene expression pattern.
- These revealed types of neurons and pathways not found in embryonic mice before.

Conclusion

- Entropy-based information measures cannot distinguish all causal dynamics.
- Ising-like interactions can offer higher resolution.
- Uniquely identify causal dynamics & logic gates.
- Möbius inversions capture higher-order structure:
  - (Pointwise) mutual information, Ising interactions are inversions on Boolean algebra
  - All have meaningful duals.
  - Other lattices:
  - categorical Ising-like interactions.
  - PID: Möbius inversion on redundancy lattice.
- Higher-order interactions are important, and estimable.

References

- A Jansma, *Higher-Order Interactions and Their Duals Reveal Synergy and Logical Dependence beyond Shannon-Information* Entropy 25 (4), 648, 2023
- SV Beentjes & A Khamseh, *Higher-order interactions in statistical physics and machine learning* Physical Review E, 102(5), 2020
- ET Jaynes, *Information theory and statistical mechanics*, Physical Review 106 (4), 620, 1957
- RG James & JP Crutchfield, *Multivariate dependence beyond Shannon information* Entropy 19, 531, 2017
- RP Stanley, *Enumerative combinatorics*, Cambridge studies in advanced mathematics, 2011