The information theory of higher-order interactions

From surprisal to Ising interactions

Abel Jansma @Abelaer

Demics23, June 2023

University of Edinburgh
Chris Ponting @ Institute of Genetics and Cancer
Luigi Del Debbio @ Higgs Centre for Theoretical Physics
Ava Khamseh @ Informatics

Outline

- **Goal: quantify higher-order structure.**
  - Information theory: *Entropy/MI*
  - Partial information decomposition
  - 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]

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

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

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

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

Model-free interactions

- What do we **really** mean when we say interaction? (Beentjes & Khamseh, 2020)
- A change in tendency to be on/off when another variable is on/off.

Tendency to be on, or *1-point interaction*:

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

2-point interaction:

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

Model-free interactions

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

$$I_{ijk} = \frac{\partial I_{ij}}{\partial X_k}\Big|_{\underline{X}=0} = \frac{\partial^3 \log p(X)}{\partial X_k \partial X_j \partial X_i}\Big|_{\underline{X}=0} \qquad \underline{X} = X \setminus \{X_i, X_j, X_k\}$$

- And so on.
- When the X_i are binary, the derivatives are just differences:

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

Model-free interactions

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

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

2-point:

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

3-point:

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

- **Model-free estimator**: sample means!
- 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!

$$E(X) = -\sum_i h_i X_i - \sum_{i,j} J_{ij} X_i X_j - \sum_{i,j,k} J_{ijk} X_i X_j X_k - \dots$$

$$\begin{aligned} I_{ijk} &= \frac{\partial^3 \log p(X)}{\partial X_k \partial X_j \partial X_i}\Big|_{\underline{X}=0} \\ &= -\frac{\partial^3 E(X)}{\partial X_k \partial X_j \partial X_i}\Big|_{\underline{X}=0} \\ &= J_{ijk} \\ &= \log \frac{p_{111} p_{100} p_{010} p_{001}}{p_{000} p_{011} p_{101} p_{110}} \approx \log \frac{\hat{n}_{111} \hat{n}_{100} \hat{n}_{010} \hat{n}_{001}}{\hat{n}_{000} \hat{n}_{011} \hat{n}_{101} \hat{n}_{110}} \end{aligned}$$

- $\hat{n}_{abc}$ is the number of samples that look like $(0, \dots, 0, a, b, c, 0, \dots, 0)$

Model-free interactions

- Surprisal of a state 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:

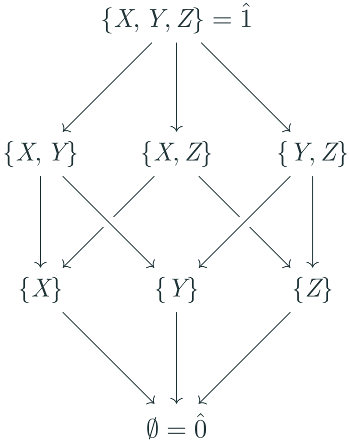
$$
\begin{aligned}
MI(X, Y) &= H(X) - H(X \mid Y) \\
&= H(X) + H(Y) - H(X, Y) \\
MI(X, Y, Z) &= MI(X, Y) - MI(X, Y \mid Z) \\
&= H(X) + H(Y) + H(Z) - H(X, Y) - H(X, Z) - H(Y, Z) + H(X, Y, Z)
\end{aligned}
$$

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

Möbius function

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

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



Möbius inversion

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:

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

- On Boolean algebra (hypercube): $\mu(x, y) = (-1)^{|x|-|y|}$

  $\implies$ Möbius inversions on Boolean algebras are **sign-alternating sums**.

Möbius inversion

- 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**:

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

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

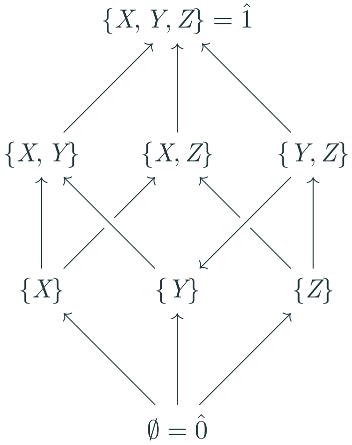
$$I(\tau; T) = \sum_{\eta \leq \tau} (-1)^{|\eta|-|\tau|} \log p(\eta = 1, T \setminus \eta = 0)$$

Dual quantities

- If $P = (S, \leq)$ is a lattice, then $P^{\text{op}} = (S, \preceq)$ (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 a generalisation of conditional entropy/differential 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:

$$
\begin{array}{ccccc}
MI^*(R) = MI(T \setminus R \mid R) & \xleftarrow{\;M_{Pop}\;} & H(R) & \xmapsto{\;M_P\;} & MI(R) \\
\Big\uparrow \mathbb{E} & & \Big\uparrow \mathbb{E} & & \Big\uparrow \mathbb{E} \\
\mathrm{pmi}^*(R = r) & \xleftarrow{\;M_{Pop}\;} & \log p(R = r) & \xmapsto{\;M_P\;} & \mathrm{pmi}(R = r) \\
\Big\downarrow eval_T & & \Big\downarrow eval_T & & \Big\downarrow eval_T \\
I^*(R;\, T) & \xleftarrow{\;M_{Pop}\;} & \log p(R = 1;\, T = 0) & \xmapsto{\;M_P\;} & I(R;\, T)
\end{array}
$$

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

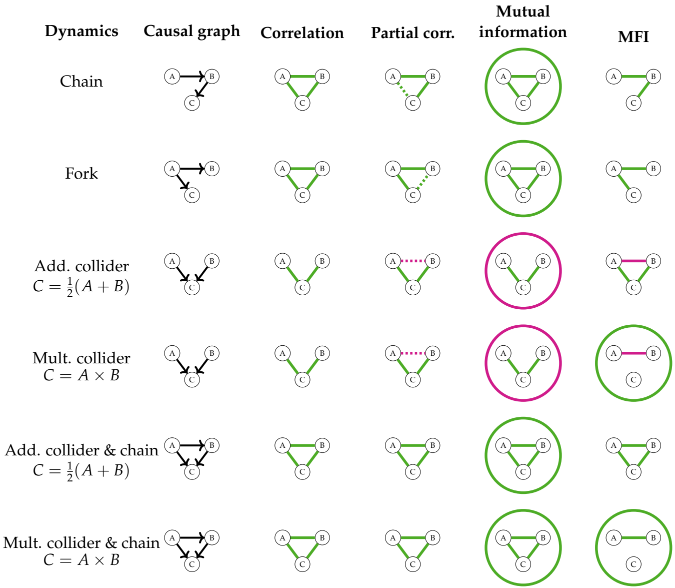
- XNOR gate!
- (XOR is maximally negative)

Results: Synergy in logic gates

- 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.
- $\boldsymbol{J}_{\boldsymbol{A}}^{\text{XNOR}} > \boldsymbol{J}_{\boldsymbol{A}}^{\text{NOR}} > \boldsymbol{J}_{\boldsymbol{A}}^{\text{AND}}$.
- Ordered by synergistic content.
- (holds for even higher-orders as well)

$\mathcal{G}$	$\boldsymbol{I_{ABC}}$	$\boldsymbol{MI_{ABC}}$	$\boldsymbol{J_A}$
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: **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

Teaser: model-free interactions on real data

- Samples that look like $(0, \ldots, 0, a, b, c, 0, \ldots, 0)$ can be rare.
- Estimation becomes tractable using Markov blankets.
- In my thesis, I calculated MFIs in gene expression data.
- 1000 genes, 20k cells
- Interactions at up to seventh order.
- These revealed types of neurons 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.
- The different notions of higher-order structure are all based on Möbius inversions:
    - (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.
- Möbius inversions capture different notions of higher-order structure.
- Ising interactions exactly disentangle different orders of dependencies, at the cost of an operational interpretation.

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