Singular learning, relative information and the dual numbers

Shaowei Lin
Topos Institute
(with Chris Hillar)

20241016
IPAM Mathematics of Intelligence Program
Theory and Practice of Deep Learning Workshop

Part I

Singular Learning

Relative Information

- Given probability distributions q and p on a *finite* set X, the *relative information* (Kullback-Leibler divergence, relative entropy) from p to q is

$$I_{q\|p}(X) = \sum_{x \in X} q(x) \log \frac{q(x)}{p(x)}.$$

- Given probability densities q and p on an *uncountably infinite* set X, the relative information is

$$I_{q\|p}(X) = \int q(x) \log \frac{q(x)}{p(x)} dx.$$

- Well-defined only when $p(x) = 0$ implies $q(x) = 0$ for all x (*absolute continuity* $q \ll p$).

Relative Information

- More generally, suppose q, p are measures with the same total measure. If $q \ll p$, let dq/dp be the Radon-Nikodym derivative. We define the *relative information* from p to q to be

$$I_{q\|p} = \int dq \, \log \frac{dq}{dp}.$$

- Think of q as the *reference* or *true* distribution, and we want to know the distance of a *model* distribution p to the truth. This distance is not symmetric, i.e. $I_{q\|p}(X) \neq I_{p\|q}(X)$.

- Relative information is well-defined for large classes of statistical models. Entropy on the other hand is often ill-defined, except in special cases where the entropy of $r(x)$ can in fact be defined as the relative information from the copy $p(x, y) = r(x)r(y)$ to the diagonal $p(x, x) = r(x)$ where $p(x, y) = 0$ if x, y distinct. To remind myself that information in a distribution should always be measured relative to another, I use the mantra

Information is relative!

Stochastic Complexity

Let $\{p(\cdot|\omega), \omega \in \Omega\}$ be a parametric model (a family of distributions) on X.
Let $\varphi(\omega)$ be a prior on the parameter space Ω. Let q be the true distribution of X.
Suppose we observe data $x_{[n]} = (x_1, \ldots, x_n) \in X^n$.

$$\text{Marginal likelihood} \qquad Z_n = \int_\Omega \prod_i p(x_i|\omega)\, \varphi(\omega) d\omega$$

$$\text{Empirical entropy} \qquad S_n = -\frac{1}{n} \sum_i \log q(x_i)$$

$$\text{Relative information} \qquad I(w) = \int q(x) \log \frac{q(x)}{p(x|\omega)} dx$$

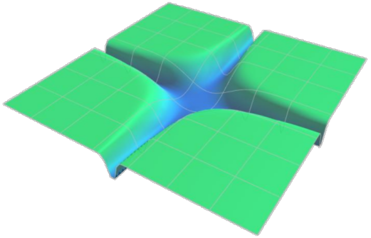
Theorem (Convergence of stochastic complexity - Watanabe)

The *stochastic complexity* has an asymptotic expansion (as $n \to \infty$)

$$-\log Z_n = nS_n + \lambda_q \log n - (\theta_q - 1) \log\log n + O_p(1)$$

where λ_q is the *real log canonical threshold* of relative information $I(w)$ over Ω with respect to $\varphi(\omega)$, and θ_q its multiplicity. For regular models, this is the Bayesian Information Criterion.

Real Log Canonical Threshold (RLCT)



- Let $f(\omega)$ be a function with $f(0) = 0$. Let Ω be a sufficiently small neighborhood of the origin.
- Asymptotically as $\varepsilon \to 0$, the volume $V(\varepsilon)$ of the tubular neighborhood approaches

$$V(\varepsilon) = \int_{\{\omega \in \Omega \,|\, f(\omega) < \varepsilon\}} d\omega \quad \rightarrow \quad C\varepsilon^{\lambda}(-\log \varepsilon)^{\theta - 1}.$$

- Using *resolution of singularities*, we can prove that λ is a positive rational number (the *real log canonical threshold* at the origin) of $f(\omega)$, and θ is a positive integer (its *multiplicity*).
- **Example.** When $f(\omega)$ is the sqr distance to a smooth mfd of codim d, then $\lambda = d/2, \theta = 1$.

Real Log Canonical Threshold of an Ideal

Theorem (RLCT of Ideal)

Suppose sets $\{f_1, \ldots, f_m\}$ and $\{g_1, \ldots, g_n\}$ of polynomials in $\mathbb{R}[x_1, \ldots, x_d]$ generate the same ideal J. Assume these polynomials all vanish at the origin. Then, the RLCTs and multiplicities of the sums of squares $f = f_1^2 + \cdots + f_m^2$ and $g = g_1^2 + \cdots + g_n^2$ are identical. Define *that* to be the RLCT of the ideal J.

Theorem (Discrete RLCT - L.)

Let X be a discrete random variable and $p(\omega) = (p_1(\omega), \ldots, p_k(\omega))$ be a parametric model. Then, the RLCT of the relative information $K(w)$ from $p(\omega)$ to $p(0)$ at the origin is the RLCT of the ideal

$$\langle\, p_1(\omega) - p_1(0), \ldots, p_k(\omega) - p_k(0) \,\rangle.$$

Theorem (Gaussian RLCT - L.)

Let X be a gaussian random variable with zero mean and $\Sigma(\omega) = (\Sigma_{11}(\omega), \Sigma_{12}(\omega), \ldots, \Sigma_{kk}(\omega))$ be the covariance matrix parametrized. Then, the RLCT of the relative information $K(w)$ from $\mathcal{N}(0; \Sigma(\omega))$ to $\mathcal{N}(0; \Sigma(0))$ at the origin is the RLCT of the ideal

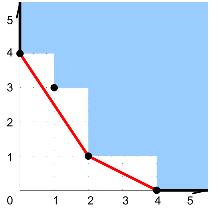
$$\langle\, \Sigma_{11}(\omega) - \Sigma_{11}(0), \ldots, \Sigma_{kk}(\omega) - \Sigma_{kk}(0) \,\rangle.$$

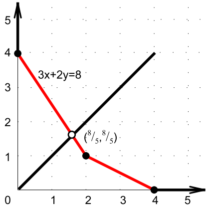
Newton Polyhedra

- Given an ideal J, for each monomial ω^α of each polynomial $f \in J$, consider the orthant $\alpha + \mathbb{R}^d_{\geq 0}$. Take the convex hull of the union of all these orthants, and call it the *Newton polyhedron* $\mathcal{P}(J)$ of J.

- Given a vector $\tau \in \mathbb{Z}^d_{\geq 0}$, define the
  1. τ-distance $l_\tau = \min\{t : t\tau \in \mathcal{P}(J)\}$.
  2. multiplicity $\theta_\tau =$ codim of face of $\mathcal{P}(J)$ at this intersection.

- e.g. $J = \langle x^4 + x^2y, xy^3 + y^4 \rangle$ and $\tau = (1, 1)$. Then $l_\tau = 8/5$ and $\theta_\tau = 1$.





Theorem (Upper bound for RLCT - L.)

The RLCT of an ideal J is bounded above by $(1, 1)$-dist and mult, with equality when J is monomial.

Bayesian Inference

- Let the *belief* on model parameters be given initially by the *prior* $p(w)$.
- Suppose we observe data $x_{[n]} = (x_1, \ldots, x_n) \in X^n$.
- We update our belief to the *posterior*

$$p(w|x_{[n]}) = \frac{p(x_{[n]}|w)p(w)}{p(x_{[n]})} = \frac{p(x_{[n]}|w)p(w)}{\int p(x_{[n]}|w)p(w)dw}.$$

- We infer new data points using the *predictive distribution*

$$p^*(x) := p(x|x_{[n]}) = \int p(x|w)p(w|x_{[n]})dw.$$

Generalization Error

- *Generalization error* G_n of Bayesian inference is the relative information from predictive distribution $p^*(X)$ to the true distribution $q(x)$.

$$G_n := I_{q||p^*}(X) = \sum_x q(x) \log \frac{q(x)}{p^*(x)}$$

- Let λ be the real log canonical threshold of the relative information

$$K(w) = \sum_x q(x) \log \frac{q(x)}{p(x|\omega)}.$$

Theorem (Watanabe[1])

$$\mathbb{E}[G_n] = \frac{\lambda}{n} + O(\frac{1}{n})$$

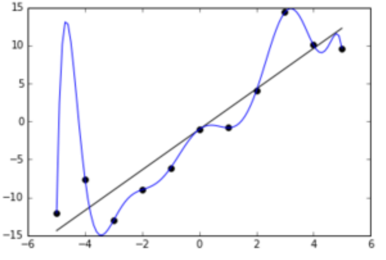
[1] Watanabe, Sumio. Algebraic geometry and statistical learning theory. Vol. 25. Cambridge university press, 2009.

Maximum Likelihood

- Likelihood of data $L_n(\omega) = \prod_i p(x_i|\omega)$
  Log-likelihood of data $\ell_n(\omega) = \log L_n(\omega) = \sum_i \log p(x_i|\omega)$

- Maximum likelihood estimate $\hat{\omega} = \arg\max_\omega \ell_n(\omega)$
  Optimize using gradient ascent with $\dot{\ell}_n(\omega) = \sum_i \frac{\partial}{\partial \omega} \log p(x_i|\omega)$.

- **Problem.** Overfitting the data. Perhaps we have been doing it wrong.



Stochastic Gradient Descent

- Suppose we could minimize the relative information (despite not knowing q).

$$K(w) := \sum_x q(x) \log \frac{q(x)}{p(x|\omega)}.$$

- Optimize using gradient descent with

$$\dot{K}(\omega) = -\sum_x q(x) \frac{\partial}{\partial \omega} \log p(x|\omega).$$

- Estimate the gradient by sampling x from q (or data $x_{[n]}$). Note similarity to $\dot{\ell}_n(\omega)$.

$$\widehat{\dot{K}}(\omega) = -\frac{\partial}{\partial \omega} \log p(x|\omega)$$

- **Advantage.** Popular technique in deep learning. Tends to overfit less. Double descent?

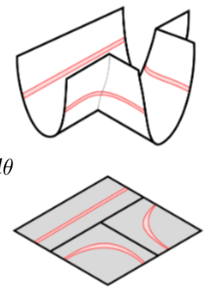
State Densities, Partition Functions and Zeta Functions!

Density of states

$$v(E) = \int \delta(I(\theta) - E)d\theta$$

Two-sided Laplace transform

Mellin transform



$$Z(N) = \int v(E)e^{-NE}dE = \int e^{-NI(\theta)}d\theta$$

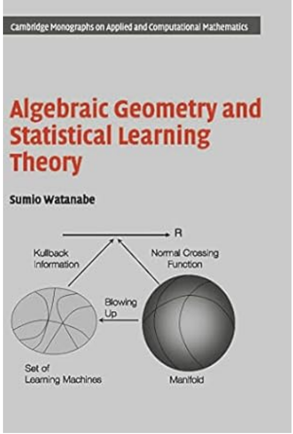
Partition function

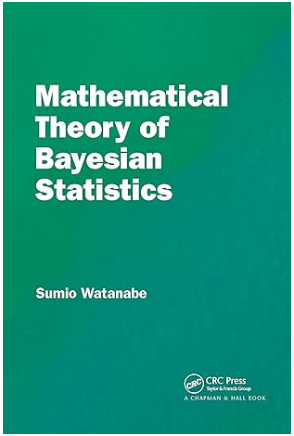
$$\zeta(z) = \int v(E)E^{z}dE = \int I(\theta)^{z}d\theta$$

Zeta function

[2]Jesse Hoogland, "Physics I: The Thermodynamics of Learning",Singular Learning Theory and Alignment Summit 2023.

More Resources





https://sites.google.com/view/sumiowatanabe/home
https://shaoweilin.github.io/singular/
https://singularlearningtheory.com/
https://timaeus.co/

Part II

Relative Information

Conditional Relative Information

- Consider joint probabilities $q(y, x)$ for $(y, x) \in Y \times X$. Conditional probabilities are $q(y|x) = q(y, x)/q(x)$ when $q(x) = \sum_y q(y, x) \neq 0$.

- Given distributions q, p on $Y \times X$, the *conditional* relative information from p to q is

$$I_{q\|p}(Y|X) = \sum_{x \in X} q(x) \sum_{y \in Y} q(y|x) \log \frac{q(y|x)}{p(y|x)}.$$

- Important concept for variational inference, expectation-maximization algorithm.

Conditional Relative Information

- More generally, given a discrete measure q on $Y \times X$, define $q(x) := \sum_y q(y,x)$ and $q(y|x) := q(y,x)/q(x)$. Let $T_q := \sum_{y,x} q(y,x)$ denote the total measure.
- Given measures q, p on $Y \times X$ such that $T_p = T_q$, the conditional relative information is

$$I_{q\|p}(Y|X) = \sum_{x \in X} q(x) \sum_{y \in Y} q(y|x) \log \frac{q(y|x)}{p(y|x)}.$$

- Normalizing $I_{q\|p}(Y|X)$ by the total measure T_q gives the statistical relative information.

Chain Rule

Theorem (Chain Rule)

$$I_{q\|p}(Y, X) = I_{q\|p}(Y|X) + I_{q\|p}(X)$$

Proof.

$$\begin{aligned}
I_{q\|p}(Y, X) &= \sum_{x,y} q(y, x) \log \frac{q(y, x)}{p(y, x)} \\
&= \sum_{x,y} q(y|x)q(x) \log \frac{q(y|x)q(x)}{p(y|x)p(x)} \\
&= \sum_{x,y} q(y|x)q(x) \log \frac{q(y|x)}{p(y|x)} + \sum_{x,y} q(y|x)q(x) \log \frac{q(x)}{p(x)} = I_{q\|p}(Y|X) + I_{q\|p}(X)
\end{aligned}$$

$\square$

Sums and Products

- Suppose we have a measure p on X and a measure q on Y.
- The sum $p + q$ is the measure on the disjoint union $X + Y$ where $(p + q)(x) = p(x)$ if $x \in X$, and $(p + q)(y) = q(y)$ if $y \in Y$.
- The product $p \times q$ is the measure on the Cartesian product $X \times Y$ where $(p \times q)(x, y) = p(x)q(y)$.
- Total measures satisfy the sum and product rules.

$$T_{p+q} = T_p + T_q$$

$$T_{p\times q} = T_p \times T_q$$

Sums and Products

- For relative information, we also have sum and product rules.
- For each $i \in \{1, 2\}$, let q_i, p_i be discrete measures on $Y_i \times X_i$ with $T_{q_i} = T_{p_i}$.

Theorem (Sum Rule)

$$I_{(q_1+q_2)\|(p_1+p_2)}(Y_1 + Y_2|X_1 + X_2) = I_{q_1\|p_1}(Y_1|X_1) + I_{q_2\|p_2}(Y_2|X_2)$$

Theorem (Product Rule)

$$I_{(q_1\times q_2)\|(p_1\times p_2)}(Y_1 \times Y_2|X_1 \times X_2) = T_{q_2} \cdot I_{q_1\|p_1}(Y_1|X_1) + T_{q_1} \cdot I_{q_2\|p_2}(Y_2|X_2)$$

Axiomatization of Relative Information

- We see that relative information satisfies the chain, sum and product rules.

- Under appropriate conditions, the only functions on probabilities that satisfy those rules are scalar multiples of relative information. There are similar axiomatization results for classical and quantum entropy. See papers below for more information.

  - Baez, Fritz, Leinster. "A characterization of entropy in terms of information loss." Entropy 13(11), 2011.
  - Baez, Fritz. "A Bayesian characterization of relative entropy." arXiv:1402.3067, 2014.
  - Baudot, Bennequin. "The homological nature of entropy." Entropy 17(5), 2015.
  - Vigneaux. "Information structures and their cohomology." arXiv:1709.07807, 2017.
  - Bradley. "Entropy as a topological operad derivation." Entropy 23(9), 2021.

Part III

Dual Numbers

Dual Numbers

- The rig (semiring) of *duals* is $\mathcal{R} = \mathbb{R}_{\geq 0}[\varepsilon]/\langle \varepsilon^2 \rangle$, where ε is an infinitesimal with $\varepsilon^2 = 0$. Denote addition by $+$ and multiplication by $\times$.
- We shall think of the rig of duals as a *category* **R**, where
  - the nonnegative reals $a \in \mathbb{R}_{\geq 0}$ are *objects*;
  - the duals $a + b\varepsilon \in \mathcal{R}$ are *morphisms* from a to itself, i.e. loops;
  - the morphisms *compose* by tangent addition $(a + b\varepsilon) \circ (a + c\varepsilon) = a + (b + c)\varepsilon$;
  - the dual $a + 0\varepsilon \in \mathcal{R}$ is the *identity* morphism from a to itself.
- Addition $+$ and multiplication $\times$ of the duals give *monoidal* structures on **R**.
  - $(a + b\varepsilon) + (c + d\varepsilon)$ is the morphism $(a + c) + (b + d)\varepsilon$ from the object $a + c$ to itself.
  - $(a + b\varepsilon) \times (c + d\varepsilon)$ is the morphism $(ac) + (ad + bc)\varepsilon$ from the object ac to itself.
- The category **R** of duals is a *rig* category.

Information Posets

- An *information poset* is a category where
  - the objects are finite sets (measurable spaces);
  - the morphisms are surjections (measurable surjections);
  - there is at most one morphism between any two objects.
  - there is a *terminal* object, a one-element set $*$.
- Disjoint union $+$ and Cartesian product $\times$ of sets give *monoidal* structures.
  - Given $f : A \to B$ and $g : C \to D$, we have $f + g : A + B \to C + D$.
  - Given $f : A \to B$ and $g : C \to D$, we have $f \times g : A \times B \to C \times D$.
- Information posets are *rig* categories.

Measure Functors

- Let **FinMeas** be the category where
  - the objects (X, μ_X) are *finite* sets equipped with a *measure*;
  - the morphisms $(Y, \mu_Y) \to (X, \mu_X)$ are measure-preserving maps, i.e. the underlying set map $f : Y \to X$ satisfies $\mu_X(x) = \mu_Y(f^{-1}(x))$.
- Fix an information poset $\mathbf{P}$. A functor $q : \mathbf{P} \to \mathbf{FinMeas}$ is a *measure functor* if it associates
  - X in $\mathbf{P}$ to some (X, q_X) in **FinMeas** where the underlying set is X;
  - $f : Y \to X$ in $\mathbf{P}$ to some $(Y, q_Y) \to (X, q_X)$ in **FinMeas** where the underlying set map is f.
  - sums $f_1 + f_2 : X_1 + X_2 \to Y_1 + Y_2$ to sums $(X_1 + X_2, \mu_{X_1} + \mu_{X_2}) \to (Y_1 + Y_2, \mu_{Y_1} + \mu_{Y_2})$.
  - products $f_1 \times f_2 : X_1 \times X_2 \to Y_1 \times Y_2$ to products $(X_1 \times X_2, \mu_{X_1}\mu_{X_2}) \to (Y_1 \times Y_2, \mu_{Y_1}\mu_{Y_2})$.
- Given a measure functor $q : \mathbf{P} \to \mathbf{FinMeas}$ and a surjection $f : Y \to X$, we define for all $y \in Y$ and $x = f(y) \in X$ with $q_X(x) \neq 0$, the *conditional probability*

$$q_f(y|x) = q_Y(y)/q_X(x).$$

Relative Information as a Functor

- Fix an information poset $\mathbf{P}$. Assume that the measure functors $q, p : \mathbf{P} \to \mathbf{FinMeas}$ have the same total measure, and that $q \ll p$.
- For each object X in $\mathbf{P}$, denote the total measure by

$$T_q(X) = \sum_{x \in X} q_X(x);$$

- For each surjection $f : Y \to X$ in $\mathbf{P}$, denote the relative information by

$$I_{q\|p}(f) = \sum_{x \in X} q_X(x) \sum_{y \in f^{-1}(x)} q_f(y|x) \log \frac{q_f(y|x)}{p_f(y|x)}$$

Theorem

Let $F_{q\|p} : \mathbf{P} \to \mathbf{R}$ be the mapping that associates each surjection $f : Y \to X$ in $\mathbf{P}$ to the dual number $T_q(X) + I_{q\|p}(f)\varepsilon$ in $\mathbf{R}$. Then $F_{q\|p}$ is a *rig monoidal functor*.

Relative Information as a Functor

Proof Outline

Claims about total measure.

- Check that $F_{q\|p}$ maps surjections $f : Y \to X$ in $\mathbf{P}$ to loops $a \to a$ in $\mathbf{R}$, i.e.

$$T_q(Y) = T_q(X).$$

- Check that $F_{q\|p}$ maps disjoint unions of objects in $\mathbf{P}$ to sums of reals in $\mathbf{R}$, i.e.

$$T_q(X_1 + X_2) = T_q(X_1) + T_q(X_2).$$

- Check that $F_{q\|p}$ maps Cartesian products of objects in $\mathbf{P}$ to products of reals in $\mathbf{R}$, i.e.

$$T_q(X_1 \times X_2) = T_q(X_1)\, T_q(X_2).$$

Indeed, the first follows because $T_q(Y)$ and $T_q(X)$ are total measures and f is measure-preserving. The second and third claims follow from the sum rule and product rule for total measure.

Relative Information as a Functor

Proof Outline

Claims about relative information.

- Check that $F_{q\|p}$ maps compositions in **P** to tangent sums in **R**, i.e.

$$I_{q\|p}(f \circ g) = I_{q\|p}(f) + I_{q\|p}(g).$$

- Check that $F_{q\|p}$ maps disjoint unions of morphisms in **P** to sums of duals in **R**, i.e.

$$I_{q\|p}(f_1 + f_2) = I_{q\|p}(f_1) + I_{q\|p}(f_2).$$

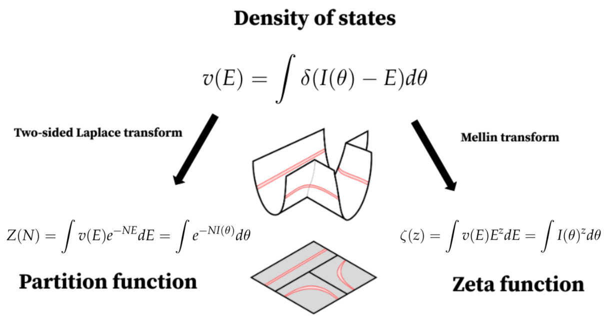
- Check that $F_{q\|p}$ maps Cartesian products in **P** to products in **R**, i.e.

$$I_{q\|p}(f_1 \times f_2) = T_q(X_2) \cdot I_{q\|p}(f_1) + T_q(X_1) \cdot I_{q\|p}(f_2).$$

Indeed, the claims follow from the chain, sum and product rules for relative information.

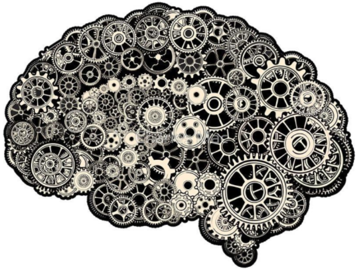
Why Relative Information?

- Information is relative! Information is energy!
- Beautiful algebra, geometry and combinatorics!
- Generalized relative information as rig monoidal functors, as cohomology.
- It from bit! [3]



[3] Wheeler, J.A. (1989). Information, physics, quantum: the search for links. Int Symp on Foundations of Quantum Mechanics. Tokyo: pp. 354-358.
[4] Jesse Hoogland, "Physics I: The Thermodynamics of Learning",Singular Learning Theory and Alignment Summit 2023.

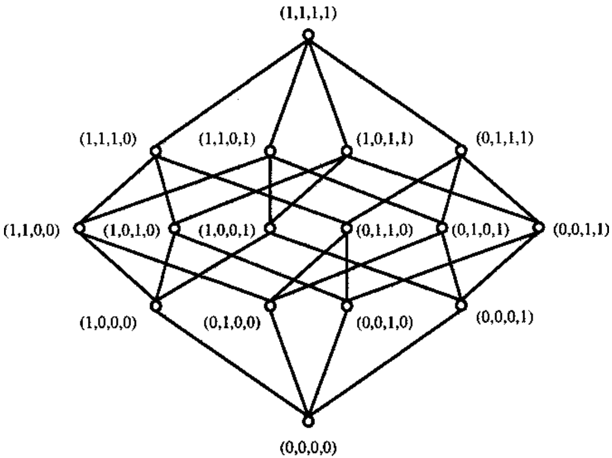
Thank you!

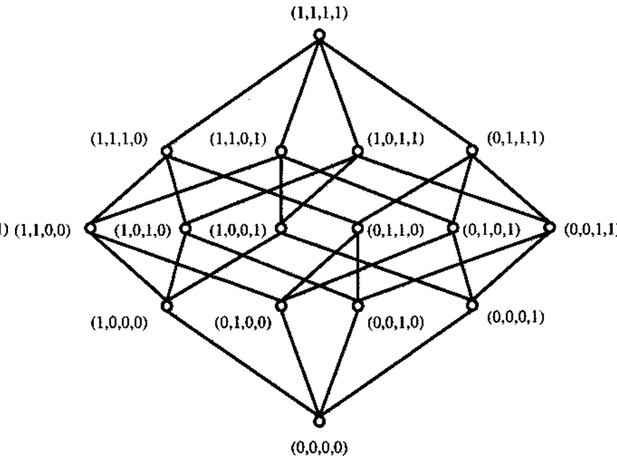


shaoweilin.github.io

SIGMA COMPLEX

- Sigma complex[5] - gluing together of sigma algebras along subalgebras.





[5] Kochen, Simon B. "A reconstruction of quantum mechanics." Quantum [Un] Speakables II: Half a Century of Bell's Theorem (2017): 201-235.

Information structures

Let $\mathbf{S}$ be a partially ordered set (poset); we see it as a category, denoting the order relation by an arrow. It is supposed to have a terminal object $\top$ and to satisfy the following property: whenever $X, Y, Z \in \operatorname{Ob}\mathbf{S}$ are such that $X \to Y$ and $X \to Z$, the categorical product $Y \wedge Z$ exists in $\mathbf{S}$. An object of X of $\mathbf{S}$ (i.e. $X \in \operatorname{Ob}\mathbf{S}$) is interpreted as an *observable*, an arrow $X \to Y$ as Y being coarser than X, and $Y \wedge Z$ as the joint measurement of Y and Z.

The category $\mathbf{S}$ is just an algebraic way of encoding the relationships between observables. The measure-theoretic "implementation" of them comes in the form of a functor $\mathcal{E} : \mathbf{S} \to \mathbf{Meas}$ that associates to each $X \in \operatorname{Ob}\mathbf{S}$ a measurable set $\mathcal{E}(X) = (E_X, \mathfrak{B}_X)$, and to each arrow $\pi : X \to Y$ in $\mathbf{S}$ a measurable *surjection* $\mathcal{E}(\pi) : \mathcal{E}(X) \to \mathcal{E}(Y)$. To be consistent with the interpretations given above, one must suppose that $E_\top \cong \{*\}$ and that $\mathcal{E}(Y \wedge Z)$ is mapped *injectively* into $\mathcal{E}(Y) \times \mathcal{E}(Z)$ by $\mathcal{E}(Y \wedge Z \to Y) \times \mathcal{E}(Y \wedge Z \to Z)$. We consider mainly two examples: the discrete case, in which E_X finite and $\mathfrak{B}_X$ the collection of its subsets, and the Euclidean case, in which E_X is a Euclidean space and $\mathfrak{B}_X$ is its Borel σ-algebra. The pair $(\mathbf{S}, \mathcal{E})$ is an *information structure*.

[6] Vigneaux, Juan Pablo. "Information cohomology of classical vector-valued observables." In Geometric Science of Information: 5th International Conference, GSI 2021, Paris, France, July 21–23, 2021, Proceedings 5, pp. 537-546. Springer International Publishing, 2021.

3.1. DEFINITION. Let $\mathbf{S}$ be a conditional meet semilattice with terminal object $\top$. We view it as a site with the trivial topology, such that every presheaf is a sheaf. For each $X \in \operatorname{Ob}\mathbf{S}$, set $\mathscr{S}_X := \{Y \in \operatorname{Ob}\mathbf{S} \mid X \to Y\}$, with the monoid structure given by the product of in $\mathbf{S}$: $(Z, Y) \mapsto ZY := Z \wedge Y$. Let $\mathscr{A}_X := \mathbb{R}[\mathscr{S}_X]$ be the corresponding monoid algebra. The contravariant functor $X \mapsto \mathscr{A}_X$ is a sheaf of rings; we denote it by $\mathscr{A}$. The pair $(\mathbf{S}, \mathscr{A})$ is a ringed site.

The category $\mathbf{Mod}(\mathscr{A})$ is abelian [Stacks Project Authors, 2018, Lemma 03DA] and has enough injective objects [Stacks Project Authors, 2018, Theorem 01DU]. For a fixed object $\mathscr{O}$ of $\mathbf{Mod}(\mathscr{A})$, the covariant functor $\operatorname{Hom}(\mathscr{O}, -)$ is always additive and left exact: the associated right derived functors are $R^n \operatorname{Hom}(\mathscr{O}, -) =: \operatorname{Ext}^n(\mathscr{O}, -)$, for $n \geq 0$.

Let $\mathbb{R}_{\mathbf{S}}(X)$ be the $\mathscr{A}_X$-module defined by the trivial action of $\mathscr{A}_X$ on the abelian group $(\mathbb{R}, +)$ (for $s \in \mathscr{S}_X$ and $r \in \mathbb{R}$, take $s \cdot r = r$). The presheaf that associates to each $X \in \operatorname{Ob}\mathbf{S}$ the module $\mathbb{R}_{\mathbf{S}}(X)$, and to each arrow the identity map is denoted $\mathbb{R}_{\mathbf{S}}$.

In Section 1.3, we have defined the *information cohomology* associated to the conditional meet semilattice $\mathbf{S}$, with coefficients in $\mathscr{F} \in \mathbf{Mod}(\mathscr{A})$, as

$$H^\bullet(\mathbf{S}, \mathscr{F}) := \operatorname{Ext}^\bullet(\mathbb{R}_{\mathbf{S}}, \mathscr{F}). \tag{29}$$

[7] Vigneaux, Juan Pablo. "Information structures and their cohomology." arXiv preprint arXiv:1709.07807 (2017).