Compositional statistical mechanics, entropy and variational inference

Gregoire Sergeant-Perthuis

LCQB Sorbonne Université

SYCO 12
Birmingham, UK 16 April, 2024

Paper arXiv:2403.16104

Statistical mechanics: motivation and problems

Statistical mechanics: motivation

- What for? $\rightsquigarrow$ modelling interaction between random variables, e.g. $X = X_1 ... X_N \in \{-1, 1\}$.
- How? $\rightsquigarrow$ Energy-based modeling
  - energy function (Hamiltonian), e.g.

$$H = X_1 \cdot X_2 + X_2 \cdot X_3 + X_3 \cdot X_4 \dots$$

  - Probability distribution: Boltzmann distribution

$$p_X \propto e^{-\beta H} \propto e^{\beta X_1 \cdot X_2} \cdot e^{\beta X_2 \cdot X_3} \cdot e^{\beta X_3 \cdot X_4} \dots \qquad \text{(Markov chain)}$$

- Relation to computer science [MM09], (key words):
  - Markov chain, Hidden Markov model (HMM)
  - Graphical models, factor graphs ...
  - Inference (maximum likelihood), Bayesian inference $\rightsquigarrow$ machine learning ($\neq$ deep learning)

Statistical mechanics: problems of interest

When there are a (countable) infinity of random variables $(X_1, \dots X_n, \dots)$ then:

- one Hamiltonian $H \to$ how to make sense of $p_X \propto e^{-\beta H}$
  - $\rightsquigarrow$ several possible distribution p (Dobrushin-Lanford-Ruelle condition)
  - $\to$ Phases, pure phases, e.g "solid, gaz, liquid"
- Characterizing the phases is a hard problem
- "Manipulating" (transforming) the phases is difficult
- For translation invariant potential:
  - $\to$ Correspondance: phases $\leftrightarrow$ tangent of a convex functional ("Helmholtz free energy") at H.
  - $\Lsh$ relates to variational principle

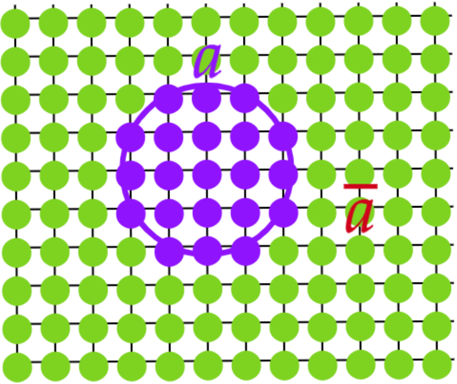
Statistical mechanics: formal definitions

Setting

- I index of (finite) random variables: $(X_i \in E_i, i \in I)$
- Global state space $\Omega = \prod_{i \in I} E_i$ denoted as E with $\sigma-$algebra $\mathscr{E}$,
- $\mathbb{P}(E)$ space of measures

- $a \subseteq I$ finite subset of I, $\mathscr{P}_f(I)$ the set of finite subsets
- $X_a \in E_a = \prod_{i \in a} E_i$ or $(E, \mathscr{E}_a)$ when seen in $(E, \mathscr{E})$
- for $b \subseteq a$, $i^a_b : E_a \to E_b$ in $\mathbf{Mes}(E_a, E_b)$

- Statistical system: collection of border conditions



- Probability kernel $p : E_{\overline{a}} \to E$,
- $p \in \mathbf{Kern}(E_{\overline{a}}, E)$
- $\forall \omega_{\overline{a}} \in E_{\overline{a}}, p_{\omega_{\overline{a}}} \in \mathbb{P}(E)$
- For $A \in \mathscr{E}$, $p(A|\omega_{\overline{a}})$ "$\cong$" $\mathbb{E}[A|\mathscr{E}_{\overline{a}}]$

Proper kernel

Definition: Proper Kernel [Geo11]

Let $\mathscr{E}_1 \subseteq \mathscr{E}$ be two σ-algebras of a set E, a kernel $p \in \mathbf{Kern}((E, \mathscr{E}_1), (E, \mathscr{E}))$ is proper if and only if, for any $A \in \mathscr{E}$, any $B \in \mathscr{E}_1$ and any $\omega \in E$,

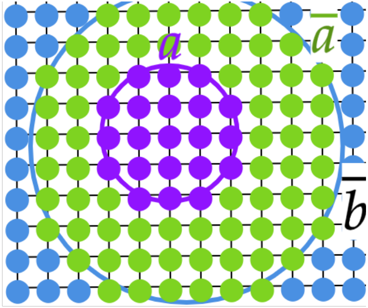
$$p(A \cap B|\omega) = p(A|\omega)1[\omega \in B]$$

For f a $\mathscr{E}_1-$measurable function,

$$p(f|\omega) = \int f(x)p(dx|\omega) = f(\omega)$$

Let $i : (E, \mathscr{E}) \to (E, \mathscr{E}_1)$ the set identity map. $p : (E, \mathscr{E}_1) \to (E, \mathscr{E})$ proper if and only if, $i \circ p = \mathrm{id}$.

Tower rule



- Tower rule: for $A \in \mathscr{E}$,

$$\mathbb{E}\left[\mathbb{E}[A|\mathscr{F}_{\overline{a}}]|\mathscr{F}_{\overline{b}}\right] = \mathbb{E}[A|\mathscr{F}_{\overline{b}}]$$

- For $\omega \in E$,

$$p_{\overline{a}} \circ p_{\overline{b}}(A|\omega_{\overline{b}}) = \int p(A|x_{\omega_{\overline{a}}})p(dx|\omega_{\overline{b}})$$

$$p_{\overline{b}}\left(p_{\overline{a}}(A|.)|\omega_{\overline{b}}\right) = p_{\overline{b}}(A|\omega_{\overline{b}})$$

Specification

Definition: Specification [Geo11]

A specification with parameter set I and state spaces $(E, \mathscr{E})$ is a collection $(\gamma_a, a \in \mathscr{P}_f(I))$ of proper kernels such that for any $a \in \mathscr{P}_f(I)$, $\gamma_a \in \mathbf{Kern}\,((E, \mathscr{E}_{\overline{a}}), (E, \mathscr{E}))$ and which satisfies that for any $a \subseteq b$, i.e $\overline{b} \subseteq \overline{a}$, any $A \in \mathscr{E}$ and $\omega \in E$,

$$\gamma_a \gamma_b(A|\omega) = \gamma_b(A|\omega)$$

Gibbs measures

Definition: Gibbs measures

Let γ be a specification with state space E, the set of probability measures,

$$\mathscr{G}(\gamma) = \{\mu \in \mathbb{P}(E) \quad : \quad \mathbb{E}_\mu(A|\mathscr{E}_{\overline{a}}) = \gamma_a(A|.)\ \mu \text{ a.s.}\}$$

is called the set of Gibbs measures of γ.

Contribution: Proposed categorical formulation

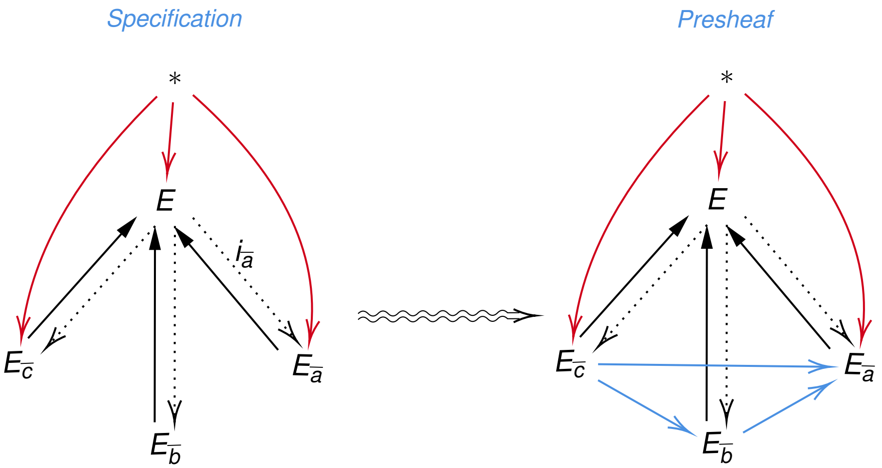
Problem:

- reference to ‘global’ $E \leftrightarrow$ difficult to “transform phases”, in fact:
  - $\rightarrow$ Hamiltonian don’t behave well with respect to maps $E \rightarrow E_1$
  - $\rightarrow$ local operation on H $(+\phi)$ are not compositional for $\mu \in \mathscr{G}(\gamma)$

What we would like:

- *“to build complex statistical systems from simple ones in a way that allows controlling and computing the phases of the associated statistical systems from the phases of the simpler one.”*
- other motivation:
  - $\rightarrow$ model statistical systems with incomplete/incompatible information on variables $+$ heterogeneity
    (related to use of sheaves in data science [SP22, Cur13, BGC$^+$22])

From classical description to categorical description



- The collection $(i_{\overline{a}}, a \in \mathscr{P}(I))$ encodes the functor of observables

Link with functors

Proposition (SP '21)

Let γ be a specification with the state space E. For any $a, b \in \mathscr{P}_f(I)$ such that $b \subseteq a$, there is a unique Markov kernel $F^a_b : E_{\overline{a}} \to E_{\overline{b}}$ such that the following diagram commutes,

$$\begin{array}{ccc} E_{\overline{a}} & \xrightarrow{\gamma_a} & E \\ {\scriptstyle F^a_b}\downarrow & \nearrow{\scriptstyle \gamma_b} & \\ E_{\overline{b}} & & \end{array} \tag{0.1}$$

i.e. such that $\gamma_b \circ F^a_b = \gamma_a$. Furthermore for any collection $a, b, c \in \mathscr{P}_f(I)$ with $a \subseteq b \subseteq c$,

$$F^b_c \circ F^a_b = F^a_c \tag{0.2}$$

Categories of measurable spaces and probability kernels

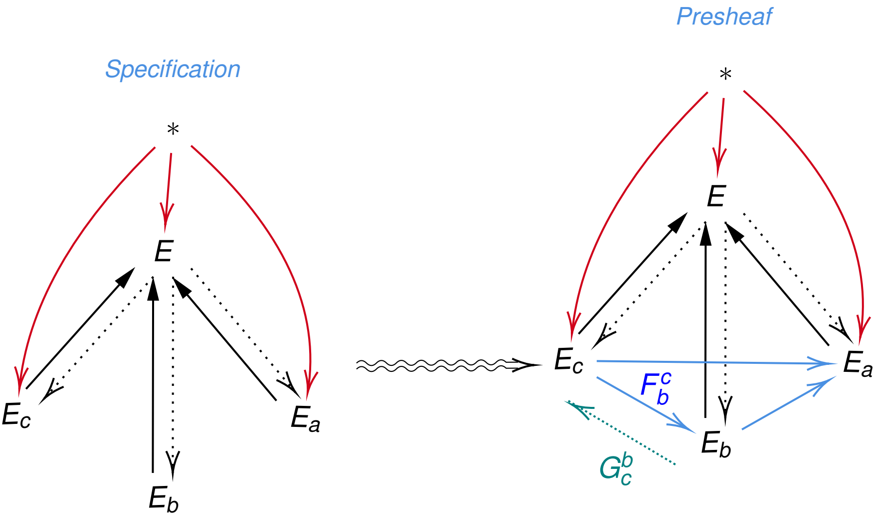
- **Mes**: Objects are measurable space, Morphisms are measurable applications
- **Kern**: Objects are measurable spaces, Morphisms are probability kernel

Proposed categorical formulation: specifications

Definition (Generalized Specification, $\mathscr{A}$-Specifications)

Let $\mathscr{A}$ be a poset, a generalized specification over $\mathscr{A}$, or simply $\mathscr{A}$-specification, is a couple (G, F) of a presheaf and a functor where $G : \mathscr{A}^{op} \to \mathbf{Mes}$ and $F : \mathscr{A} \to \mathbf{Kern}$ are such that for any $a, b \in \mathscr{A}$ with $b \leq a$,

$$G^a_b F^b_a = \text{id} \tag{0.3}$$



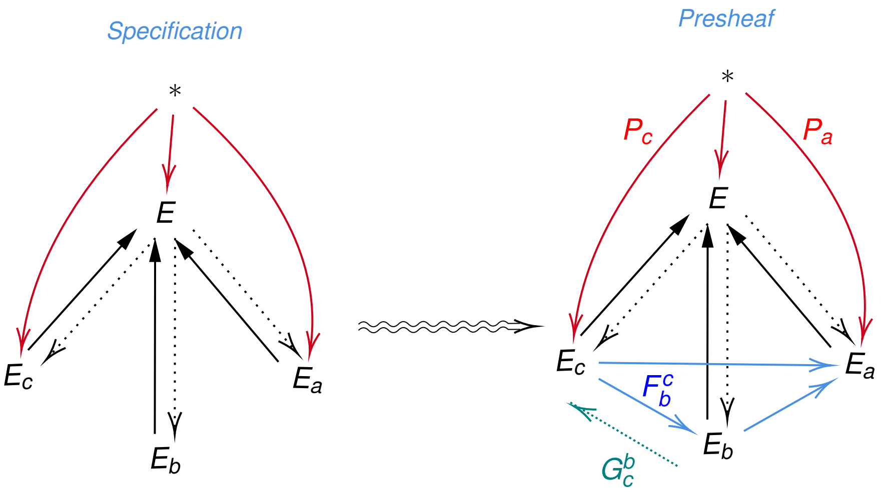
$$G_c^b F_b^c = \mathrm{id}$$

Proposed categorical formulation: Gibbs measures of a specification

Definition 8.3.7: Gibbs measures for specifications

Let $\gamma = (G, F)$ be a specification over $\mathscr{A}$, we shall call the Gibbs measures of γ the sections of F,

$$\mathscr{G}_g(\gamma) = \{P_a \in \mathbb{P}(F(a)), a \in \mathscr{A} | \quad \forall b \leq a, F_a^b P_b = P_a\}$$


Specification
Presheaf
*
E
E_c
E_b
E_a
P_c
P_a
F_b^c
G_c^b

Use homological algebra to compute phases.

- Phases are invariants: H^0
- Use resolutions to compute H^0.
- Projective and injective objects characterized in previous work (next slides)

$\implies$ Projectives extend independent random variables
Corollary 3.2 [SP20] (necessary condition harder)

Projective presheaves

Definition (Projective presheaf)

$F : \mathscr{A}^{op} \to \textbf{Vect}$ is projective if there is a collection S_a, $a \in \mathscr{A}$ such that,

$$F(a) \cong \bigoplus_{c \leq a} S_c \tag{0.4}$$

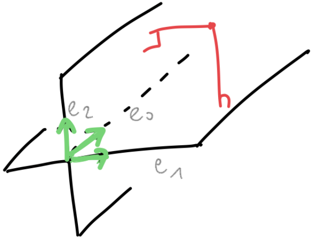
and,

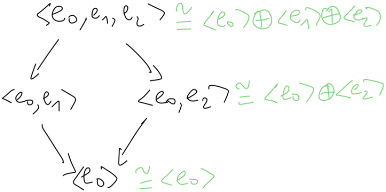
$$F_b^a \cong pr_b^a : \bigoplus_{c \leq a} S_c \to \bigoplus_{c \leq b} S_c \tag{0.5}$$

$\to$ Characterized in [SP20]: "$F_b^a \circ F_c^a = F_{b \wedge c}^a$"

Projective presheaves: example

- $V := \mathbb{R}^3$, $V_1 := \langle e_0, e_1 \rangle$, $V_2 := \langle e_0, e_2 \rangle$, $V_3 := \langle e_0 \rangle$
- Projections: orthogonal with respect to scalar product
- Collection of projections defines a projective presheaf





- Change scalar product $\rightsquigarrow$ not projective anymore.

Projective specifications

- $L^\infty(E)$, the set of bounded, real-valued, measurable functions over E
- L^∞: presheaf from **Mes** and **Kern** to the category of vector spaces **Vect**
- Consider $L^\infty \circ F : \mathscr{A}^{op} \to \textbf{Vect}$

Definition (Projective $\mathscr{A}$-specifications)

An $\mathscr{A}$-specification (G, F) is called projective when $L^\infty \circ F$ is a projective presheaf (in **Vect**).

Phases of projective presheaves

- poset $\mathscr{A}$, equivalence relation by symmetrizing the order:

$$\forall a, b \in \mathscr{A}, \; a \sim b \iff a \leq b \text{ or } b \leq a \tag{0.6}$$

- equivalence classes are connected components
- denote $\mathscr{C}(a)$ the connected component of $a \in \mathscr{A}$
- if each connected component has a minimum element, denote $\mathscr{C}_*(\mathscr{A})$ the set of these minimas. If not $\mathscr{C}_*(\mathscr{A}) := \emptyset$

Phases of projective presheaves

Theorem (SP '21)

Let $\gamma = (G, F)$ be a projective $\mathscr{A}$-specification. If at least one of the connected components of $\mathscr{A}$ does not have a minimum element, i.e. when,

$$\mathscr{C}_*(\mathscr{A}) = \emptyset \tag{0.7}$$

then,

$$\mathscr{G}_g(\gamma) = \emptyset \tag{0.8}$$

if not,

$$\mathscr{G}_g(\gamma) = \prod_{a \in \mathscr{C}_*(\mathscr{A})} \mathbb{P}(\gamma(a)) \tag{0.9}$$

Characterizing extreme Gibbs measures of $\mathscr{A}$ -specifications

$\hookrightarrow$ ~~towards~~ a $0-1$ law for extreme Gibbs measures of $\mathscr{A}$ -specifications

Classical setting

For simplicity of presentation: $I = \mathbb{N}$,

- $E = \prod_{i \in \mathbb{N}} E_i$
- $\mathscr{E}_{\geq k}$, σ-algebra generated by the cylinders: $i_{\geq k}^{-1}(A)$ where $A \subseteq \prod_{n \geq k} E_n$.
- tail σ-algebra: $\mathscr{E}_\infty := \bigcap_{k \in \mathbb{N}} \mathscr{E}_{\geq k}$.
- Set of Gibbs measures $\mathscr{G}(\gamma)$ is convex.

From [Geo11]):

→ Gibbs measures are fully characterized by their restriction to the tail σ-algebra $\mathscr{E}_\infty$.

→ Extreme Gibbs measures, $\mu \in \text{ext}\, \mathscr{G}_g(\gamma)$, are 'trivial' on $\mathscr{E}_\infty$, i.e. for any $A \in \mathscr{E}_\infty$,

$$\mu(A) = 0 \text{ or } 1$$

Assumptions

- Recall $\gamma = (G, F)$, $G : \mathscr{A}^{op} \to \mathbf{Mes}$, $F : \mathscr{A} \to \mathbf{Kern}$
- Assumptions: $G(a)$ are finite ($\mathscr{A}$ does not need to be finite) and a "positivity condition" on F.
  - ↼ Assumption (probably) not required for weaker version of $0-1$ law.

Tail σ-algebra for $\mathscr{A}$-specifications

Definition (Tail σ-algebra for $\mathscr{A}$-specifications)

Let $\gamma = (G, F)$ be a $\mathscr{A}$-specification. Pose $\sigma(G)^b_a A_b := G^{a^{-1}}_b A_b$,for $b \leq a$. The tail σ-algebra of $\mathscr{A}$-specification is defined as:

$$\lim \sigma(G) := \{(A_a \in \sigma(G(a)), a \in \mathscr{A}) | \forall a, b \in \mathscr{A}, \quad A_a = G^{a^{-1}}_b A_b\}$$

$0 - 1$ law for $\mathscr{A}$-specifications

Theorem (SP '24, Section 7 [SP24])

Let $\gamma = (G, F)$ be a specification, let $G(a)$ be finite sets for any $a \in \mathscr{A}$, let $F > 0$. $\mathscr{G}_g(\gamma)$ is a convex set. Each $\mu \in \mathscr{G}_g(\gamma)$ is uniquely determined by it's restriction to $\lim \sigma(G)$. Furthermore μ is extreme in $\mathscr{G}_g(\gamma)$ if and only if for any $A \in \lim \sigma(G)$, $\forall a \in \mathscr{A}$, $\mu_a(A_a) = 0$ or 1.

$\rightarrow$ Echoes with $0-1$ laws in a categorical setting for i.i.d sequences and Markov chains [FGP21, MP23].

$\hookrightarrow$ Full relation to be explored in future work.

Toward an entropy and variational principle for $\mathscr{A}$-specifications
$\rightsquigarrow$ Section 8 Paper arXiv:2403.16104

Thank you for your attention

Thank you for your attention!

References I

Cristian Bodnar, Francesco Di Giovanni, Benjamin Paul Chamberlain, Pietro Lio, and Michael M. Bronstein, Neural sheaf diffusion: A topological perspective on heterophily and oversmoothing in GNNs, Advances in Neural Information Processing Systems (Alice H. Oh, Alekh Agarwal, Danielle Belgrave, and Kyunghyun Cho, eds.), 2022.

D. Bennequin, O. Peltre, G. Sergeant-Perthuis, and J. P. Vigneaux, Extra-fine sheaves and interaction decompositions, September 2020, Submitted to Algebraic and Geometric Topology, arXiv:2009.12646.

Justin Curry, Sheaves, cosheaves and applications, Ph.D. thesis, The University of Pennsylvania, 2013, arXiv:1303.3255.

Tobias Fritz, Tomáš Gonda, and Paolo Perrone, De finetti's theorem in categorical probability, arXiv preprint arXiv:2105.02639 (2021).

Hans-Otto Georgii, Gibbs measures and phase transitions, De Gruyter, Berlin, New York, 2011.

Marc Mezard and Andrea Montanari, Information, physics, and computation, Oxford University Press, Inc., USA, 2009.

References II

Sean Moss and Paolo Perrone, A category-theoretic proof of the ergodic decomposition theorem, Ergodic Theory and Dynamical Systems **43** (2023), no. 12, 4166–4192.

D. Ruelle, Rigorous results in statistical mechanics, Lect. Theor. Phys. **6** (1964), 73–111.

G. Sergeant-Perthuis, Bayesian/Graphoid intersection property for factorisation spaces, March 2019, arXiv:1903.06026.

______, Intersection property and interaction decomposition, April 2019, arXiv:1904.09017.

______, Interaction decomposition for presheaves, August 2020, arXiv:2008.09029.

______, Interaction decomposition for Hilbert spaces, 2021, arXiv:2107.06444.

Grégoire Sergeant-Perthuis, Intersection property, interaction decomposition, regionalized optimization and applications, 2021, PhD thesis, 10.13140/RG.2.2.19278.38729, Link to manuscript.

References III

______, Regionalized optimization, 2022.

______, Compositional statistical mechanics, entropy and variational inference, 2024.