Lecture 5 - INFERENCE AND STATISTICAL PHYSICS

1. Notions from Statistical Physics

Stat. Physics: How can complex behavior emerge from the interaction of a large number of identical particles (e.g. water)

→ probabilistic description of the physical system in terms of elementary particles

→ extract macroscopic / deterministic facts about the system through moments / expectations under this probabilistic model. → qualitative behavior

- The Boltzmann distribution

→ N particles x_i, each belong to a configuration space $\mathcal{X}$

$$x = (x_1, \dots, x_n) \in \underbrace{\mathcal{X} \times \dots \times \mathcal{X}}_{N} =: \mathcal{X}_N$$

e.g. $x_i \in \mathcal{X} = \{+1, -1\}$ a spin $\uparrow$ or $\downarrow$

$$|\mathcal{X}| = 2, \quad |\mathcal{X}_N| = 2^N \;!$$

→ Observable: a function $x \mapsto O(x)$ on the configuration space

→ Energy: $E(x)$, a fundamental observable

$$E(x) = \sum_{i=1}^{N} E_i(x_i) \quad : \text{ non-interacting system}$$

$$E(x) = \sum_{i_1, \dots, i_K} E_{i_1, \dots, i_K}(x_{i_1}, \dots, x_{i_K}) : \text{ K-body interation system.}$$

Usually, K is small ($K=2$ or 3) in real physical system $\Rightarrow K \ll N$, source of structure.

Def: (Boltzmann distribution)
Distribution $p_\beta(x)$ on the (joint) configuration space

$$p_\beta(x) = \frac{1}{Z(\beta)} e^{-\beta E(x)}$$

- $\beta = \frac{1}{T}$ is the inverse temperature
- $Z(\beta) = \sum_{x \in \mathcal{X}} e^{-\beta E(x)}$ is the partition function

Remarks:
- This is a Gibbs distribution! (Lec. 4)
- A maximum entropy distr. :

$$\max_p H(p) \quad \text{s.t.} \quad \mathbb{E}_p[E] = y$$

note: expectations typically written as

$$\langle O \rangle_\beta = \mathbb{E}_{p_\beta}[O(x)]$$

- For continuous x, p_β is a density w.r.t. Lebesgue measure.

Role of the inverse temperature β

- Just a parameter: could change E and set $\beta = 1$
- Yet, $\beta = \frac{1}{T}$ carries physical meaning

$\rightarrow$ $\beta \to 0$ (High-temperature limit)

$$p_\beta(x) \xrightarrow[\beta \to 0]{} \frac{1}{|\mathcal{X}|} \quad \text{uniform distribution}$$

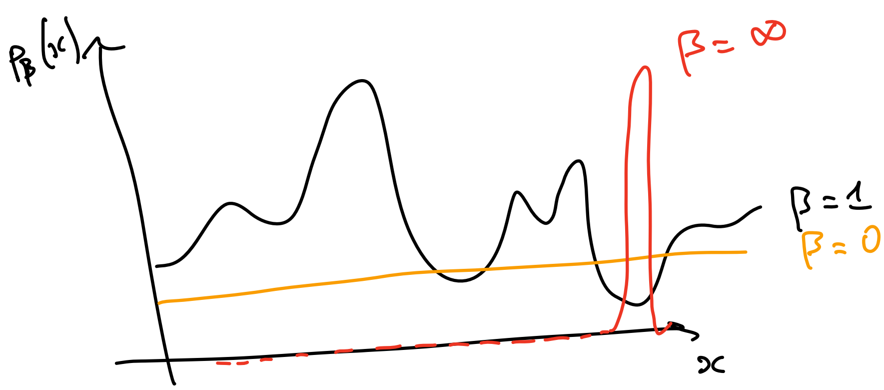
$\rightarrow$ $\beta \to \infty$ (low-temperature limit)

$$p_\beta(x) \xrightarrow[\beta \to \infty]{} \frac{1}{|\mathcal{X}_0|} \mathbb{1}\{x \in \mathcal{X}_0\}$$

$$\mathcal{X}_0 = \arg\min E = \{x_0 : E(x) \geq E(x_0) \ \forall x \in \mathcal{X}\}$$

set of Ground States (global minima)

(p_β is uniform over ground state)



Thermodynamic potentials

Quantities that summarize properties of Boltzmann dist p_β

- **Free Energy**: $F(\beta) = -\frac{1}{\beta} \log Z(\beta)$

  (also, "free entropy" $\phi(\beta) = -\beta F(\beta) = \log Z(\beta)$)

- **Internal Energy**: $U(\beta) = \frac{\partial}{\partial \beta}\left(\beta F(\beta)\right)$

- **Canonical Entropy**: $S(\beta) = \beta^2 \frac{\partial}{\partial \beta} F(\beta)$

Fact:

$$U(\beta) = \langle E(x) \rangle = \mathbb{E}_{p_\beta}[E(x)]$$

$$S(\beta) = -\sum_x p_\beta(x) \log p_\beta(x)$$

$$F(\beta) = U(\beta) - \frac{1}{\beta} S(\beta)$$

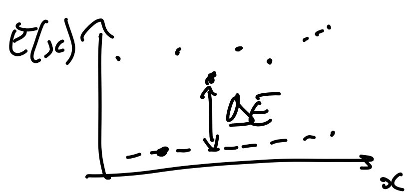
Limits:

- $\beta \to 0$:

$$F(\beta) = -\frac{1}{\beta} \log |\mathcal{X}| + \langle E(x) \rangle_0 + \Theta(\beta)$$

- $\beta \to \infty$: let $\Delta E = \min\{E(x) - E_0,\ x \notin \mathcal{X}_0\}$

  (energy gap)



$$U(\beta) = E_0 + \mathcal{O}(e^{-\beta\Delta E})$$

$$S(\beta) = \log|\mathcal{X}_0| + \mathcal{O}(e^{-\beta\Delta E})$$

$$\left[\, F(\beta) = E_0 - \frac{1}{\beta}\log|\mathcal{X}_0| + \mathcal{O}(e^{-\beta\Delta E}) \right.$$

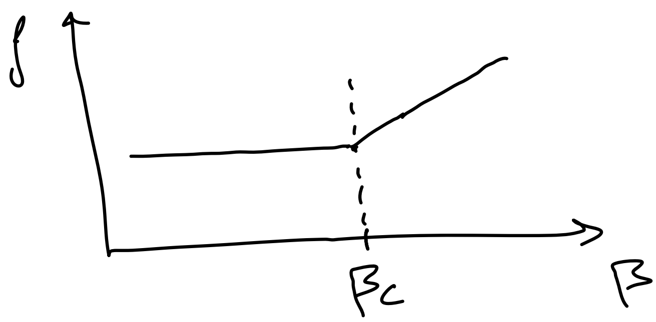
("Laplace method")

- Thermodynamic limit

$N \to \infty$ limit, describes macroscopic properties of the system.

- Thermodynamic potentials are often proportional to N

(free energy density) $\displaystyle f(\beta) = \lim_{N\to\infty} \frac{F_N(\beta)}{N}$



singularities in $f(\beta)$ are called

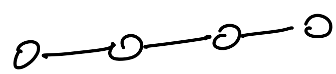
phase transitions

often correspond to qualitative changes in the system

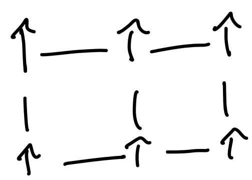
(e.g. boiling Temp of water)

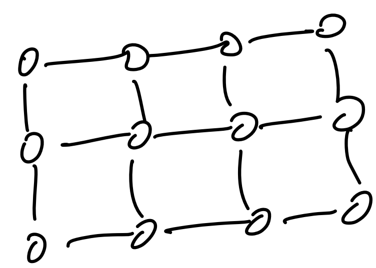
Example: Ising model

- model of magnetic materials
- molecules tend to align with magnetic field
- — also interact with each other
- Ising model represents molecules as "Ising spins" on a lattice $\mathbb{L} = \{1 \dots L\}^d$



(d=1)





(d=3)

spins $\sigma_i \in \{+1, -1\}$

Configuration $\sigma = (\sigma_1, \dots, \sigma_N)$ has energy

$\left(\mathcal{X}_N = \{\pm 1\}^N \quad N = L^d \right)$

$$E(\sigma) = - \underbrace{\sum_{i \sim j}}_{\text{edges}} \underbrace{\sigma_i \sigma_j}_{\text{pairwise interactions}} - B \sum_i \sigma_i$$

↳ external magnetic field

Typical quantity of interest: average magnetization

$$M_N(\beta, B) = \frac{1}{N} \sum_i \langle \sigma_i \rangle_\beta$$

and its limits.

→ d=1 (Ising 1925): no phase tr.

→ d=2 (1948, Onsager): phase tr.

→ d>2 : open

Ex: Ising Spin glass

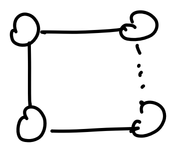
different types of interactions:

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

$J_{ij} > 0$: ferromagnetic

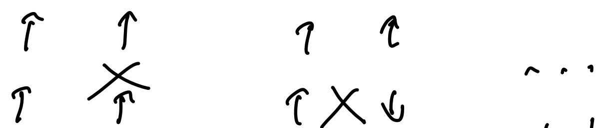
$J_{ij} < 0$: anti-ferromagnetic

→ this can lead to "frustration"



$\cdots\cdots$: $J_{ij} < 0$

—— : $J_{ij} > 0$



no configuration can satisfy all constraints!

→ much less understood than Ising models

→ Giorgio Parisi (Nobel prize today!)

"Statistical Field Theory" (1988)

$\boxed{2.}$ Inference in graphical models

$$p(x_1,\dots,x_N) = \frac{1}{Z}\prod_{c\in\mathcal{C}} \phi_c(x_c)$$

$\mathcal{C}$: cliques of graph G.

→ We are interested in queries of the form

$$p(x_S) = \sum_{x_{V\setminus S}} p(x_1,\dots,x_N)$$

(marginalization)

→ this will also allow computing conditional probabilities (by Bayes rule)

→ (reminder) This is intractable without structure (at least for $|S| \ll N$)

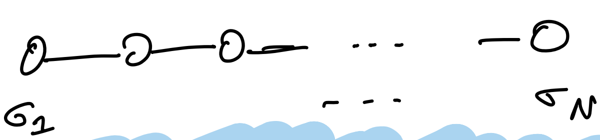
⇒ need to exploit graph structure

→ Algorithms based on message passing

Belief propagation / Sum-Product / Junction Tree
(BP)

BP on Ising chains

(1D Ising model)



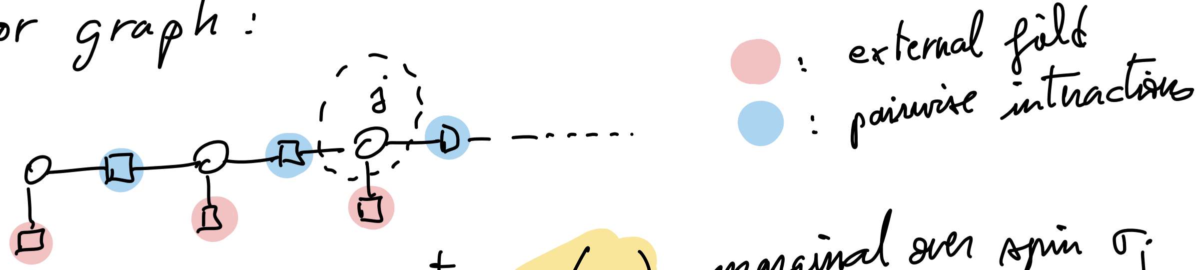
$\sigma = (\sigma_1,\dots,\sigma_N)$

$\sigma_i \in \{\pm 1\}$

$$E(\sigma) = -\sum_{i=1}^{N-1} \sigma_i\sigma_{i+1} - B\sum_{i=1}^{N}\sigma_i$$

$$P_\beta(\sigma) \propto e^{-\beta E(\sigma)}$$

- Factor graph:



$\rightarrow$ want to compute $p(\sigma_j)$, marginal over spin σ_j

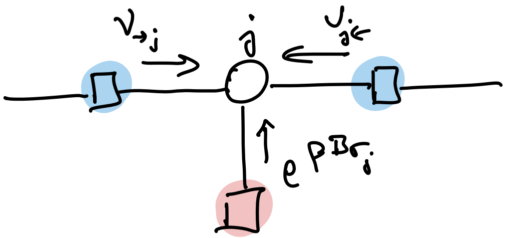
- Introduce messages $\nu_{\to j}$ and $\nu_{\leftarrow j}$

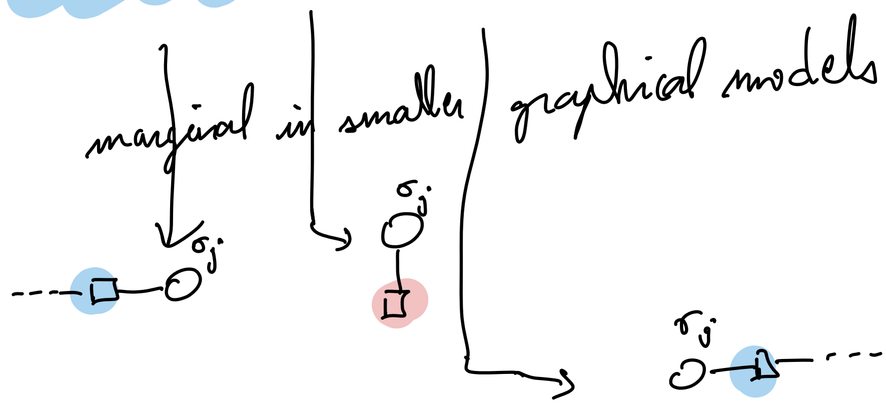
$$\nu_{\to j}(\sigma_j) = \frac{1}{Z_{\to j}} \sum_{\sigma_1,\dots,\sigma_{j-1}} \exp\left\{\beta \sum_{i=1}^{j-1} \sigma_i \sigma_{i+1} + \beta B \sum_{i=1}^{j-1} \sigma_i\right\}$$

$$\nu_{j\leftarrow}(\sigma_j) = \frac{1}{Z_{j\leftarrow}} \sum_{\sigma_{j+1},\dots,\sigma_N} \exp\left\{\beta \sum_{i=j}^{N} \sigma_i \sigma_{i+1} + \beta B \sum_{i=j+1}^{N} \sigma_i\right\}$$

We then have (check!)

$$p(\sigma_j) \propto \nu_{\to j}(\sigma_j)\, e^{\beta B \sigma_j}\, \nu_{j\leftarrow}(\sigma_j) \qquad (*)$$





- Computing messages:

Key observation:

$$\nu_{\to j}(\sigma_j) \propto \sum_{\sigma_{j-1}} \nu_{\to j-1}(\sigma_{j-1}) \exp\left(\beta \sigma_{j-1}\sigma_j + \beta B \sigma_j\right) \qquad (**)$$

$\Rightarrow$ only need sum over a <u>single</u> spin!

An instance of Dynamic Programming:

Algoritm for computing $p(\sigma_j)$

(i) initialize $\nu_{\to 1}$, $\nu_{N\leftarrow}$ to <u>uniform</u>

(ii) Compute all messages $\nu_{\to j}$, $\nu_{\leftarrow j}$ using Dynamic Programming (**)

($\nu_{\to j}$: forward, $\nu_{j\leftarrow}$: backward)

(iii) Compute desired marginals using (*)

<u>Remark</u>: . $O(N)$ time complexity, $O(1)$ space

. if we had K state $x_i \in \{1, \dots, K\}$, $O(NK^2)$ time

. What about marginals $p(\sigma_j, \sigma_k)$ $j<k$?

$\to$ sufficient to compute $p(\sigma_k | \sigma_j)$

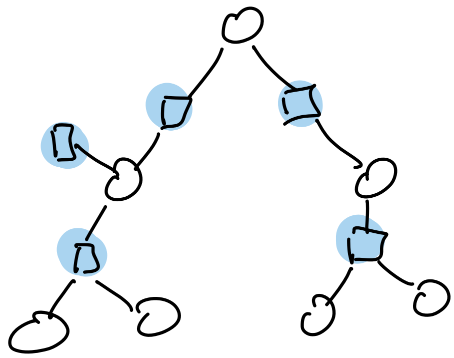
$\to$ note that $p(\sigma_{j+1}, \dots, \sigma_N | \sigma_j)$ take a similar form $\propto \exp\{\beta \sum_{i=j}^{N-1} \sigma_i \sigma_{i+1} + \beta B \sum_{i=j}^{N} \sigma_i\}$.

$\to$ computing $p(\sigma_k | \sigma_j) \Leftrightarrow$ computing marginal over σ_k in this modified model, with $\neq$ initial conditions

BP on trees (preview)

- Consider graphical models that are trees (no cycles)

$$p(x) = \frac{1}{Z} \prod_c \phi_c(x_c)$$



- In this case, BP is also <u>exact</u>

- Messages $\nu_{i\to a}$ (node → factor)

  $\nu_{a\to i}$ (factor → node)

Updates (sum-product):

$$\nu^{(t+1)}_{j\to a}(x_j) \propto \prod_{b\in N(j)\setminus\{a\}} \nu^{(t)}_{b\to j}(x_j)$$

$$\nu^{(t)}_{a\to j}(x_j) \propto \sum_{x_{N(a)\setminus\{j\}}} \phi_a(x_{N(a)}) \prod_{k\in N(a)\setminus j} \nu^{(t)}_{i\to a}(x_k)$$

Marginals: $\nu_i^{(t)}(x_i) \propto \prod_{a\in N(i)} \nu^{(t-1)}_{a\to i}(x_i)$

(Thm): After some $t > t^*$, we have $\nu_i^{(t)}(x_i) = p(x_i)$