STAT 206A: Gibbs Measures	Fall 2006

Lecture 1

Lecture date: Aug 29 — *Scribe: Madhur Tulsiani*

This lecture considers a few historical and motivating examples.

1 The Ising Model

The Ising model is used to model the spins of atoms in a physical system. A commonly studied instance is the $N \times N$ grid $G_N = (V_N, E_N)$ with each vertex representing an atom. Each atom i has a spin $\sigma_i \in \{-1, 1\}$ and hence the configuration space is $\{-1, 1\}^{V_N}$. The probability of each configuration is given by the distribution

$$\mathbf{P}[\sigma] = \frac{1}{Z} \prod_{(u,v)\in E_N} \exp(\beta \sigma_u \sigma_v)$$

where $\beta = 1/T$, is the inverse temperature, which determines the randomness of the system. Notice that

- $\beta = 0$ corresponds to the uniform distribution and σ_u, σ_v are independent for all u and v.
- For $\beta = \infty$, the measure is uniform over the all-plus and all-minus configurations, while all others are disallowed. Hence, $\sigma_u = \sigma_v$ for all u and v.

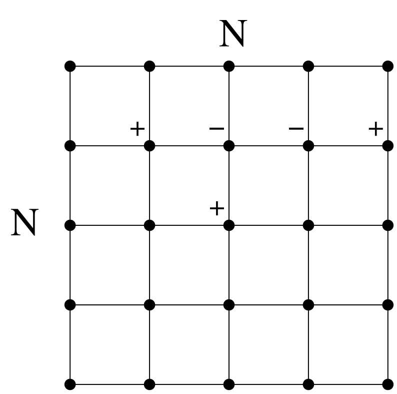

Figure 1: The Ising model

Also for the two dimensional grid, it is known that $\exists\ 0 < \beta_C < \infty$ such that,

- $\beta < \beta_C \implies \mathbf{E}[\sigma_u \sigma_v] < \exp(-c(\beta) d(u,v))$
- $\beta > \beta_C \implies \forall N, \forall u, v \;\; \mathbf{E}[\sigma_u \sigma_v] \geq \epsilon(\beta)$

where $c(\beta)$ and $\epsilon(\beta)$ are positive functions of ϵ (but independent of N). Hence, the correlation between the spins of two atoms experiences a "phase transition" from decaying exponentially with the distance, to being present over arbitrarily long distances. Finally, also notice that the above distribution has the so called "Markov Property" i.e. for every partition of V_N into disjoint sets I(inside), O(outside) and B(boundary) such that there are no edges $(u,v) \in E_N$ with $u \in I$ and $v \in O$, the configurations of I and O are independent, given a configuration for B. Formally

$$\mathbf{P}[\sigma(O) = X(O), \sigma(I) = X(I) | \sigma(B) = X(B)] =$$
$$\mathbf{P}[\sigma(O) = X(O) | \sigma(B) = X(B)] \mathbf{P}[\sigma(I) = X(I) | \sigma(B) = X(B)]$$

where $\sigma(S)$ and $X(S)$ are used to denote the restrictions of these vectors to the set S.

Exercise 1 (0.5 pts) *Prove that the above distribution satisfies the Markov property.*

Even though this is a random distribution over the configurations, the system itself is completely defined. The next example introduces randomness in the description of the system itself.

2 Spin Glasses

Spin glasses are defined on the same model as before (grid). However, the probability distribution is given by

$$\mathbf{P}[\sigma] = \frac{1}{Z} \prod_{(u,v) \in E_N} \exp(\beta J_{uv} \sigma_u \sigma_v)$$

where each $J_{uv} \in -1, 1$ is an independent random variable taking each value with probability 1/2. Hence, neighbors may be inclined towards having alike or different spins. The maximum probability configuration is given by the solution to

$$\max \sum_{(u,v) \in E_N} J_{uv} \sigma_u \sigma_v = \max \sum_{J_{uv}=1} \sigma_u \sigma_v - \sum_{J_{uv}=-1} \sigma_u \sigma_v$$

Finding such a configuration on a general graph is known as correlation clustering and is known to be NP-complete. This is an example of a general phenomena: spin glasses turn out to be much harder to analyze than Ising models. It is also an open problem whether there is a phase transition in spin glass models. Note that by the symmetry of spin-glasses it holds that $E[E_J[\sigma_u \sigma_v]] = 0$ for every $u \neq v$. However one may phrase the phase transition in terms of the typical behavior. For example one way to formula the phase transition problem is to ask if there exists β_c s.t. for all $\beta > \beta_c$ it holds that: $P[|E_J[\sigma_u \sigma_v]| > \epsilon] > \epsilon$

3 SAT formulas and Codes

A SAT formula (in CNF form) is a is a function $\psi(X)$ on a set X of variables $x_1 \dots x_n$ of of the form $c_1 \wedge c_2 \wedge \dots$, where each c_i (called a clause) and is of the form $x_{i_1} \vee \bar{x}_{i_2} \dots$. Each $x_i \in \{0, 1\}$ and $A \vee B$, $A \wedge B$ are interpreted as $max(A, B)$ and $min(A, B)$ respectively. Equivalently, we may think of the variables as being true or false and $\vee$, $\wedge$ as the OR and AND functions respectively. The problem is to find an assignment to the variables such that the formula is satisfied.

We draw the "Factor Graph" of the SAT formula as a bipartite graph with one vertex u_i on the left for each variable x_i and one vertex v_j on the right for each clause c_j. The edges are defined by the containment of a variable in a clause i.e. $(u_i, v_j) \in E$ iff c_j contains x_i or $\bar{x}_i$.

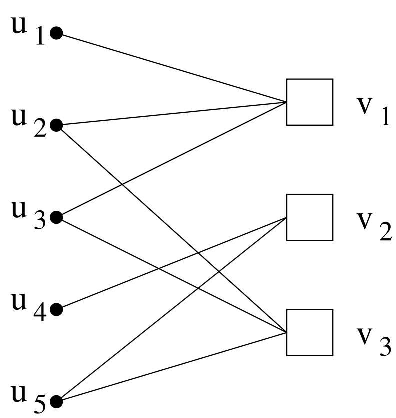

Figure 2: Factor graph for $(x_1 \vee x_2 \vee x_3) \wedge (\bar{x}_3 \vee x_4) \wedge (\bar{x}_2 \vee x_3 \vee x_5)$

One can see that the uniform distribution over all satisfying assignments to the variables satisfies a Markov property similar to the one in the Ising model i.e. if I, O and B partition the set of all variables s.t. no clause contains variables form both I and O, then

$$\mathbf{P}[X_I = y_I, X_O = y_O | X_B = y_B] = \mathbf{P}[X_I = y_I | X_B = y_B] \mathbf{P}[X_O = y_O | X_B = y_B]$$

Another example might is linear codes. These are defined by a set of linear constrains over $\mathbb{Z}_2$, say the vectors $x \in \{0, 1\}^6$, which satisfy

$$\begin{aligned} x_1 + x_2 + x_3 &= 0 \ (mod\ 2) \\ x_4 + x_5 + x_6 &= 0 \ (mod\ 2) \\ x_1 + x_4 &= 0 \ (mod\ 2) \end{aligned}$$

One may draw a factor graph for the above constraint system in the same way as for the SAT formula, replacing the clause by an equation and defining edges similarly. The uniform distribution over the set of valid assignments to the variables factorizes in the same way as in SAT. In general, we try to analyze large random objects (SAT formulas, codes etc.) and look at distributions which factorize over graphs i.e. each vertex is correlated only with the vertices to which it is connected.

4 Role of Trees

Trees are particularly interesting in studying these kinds of measures as it is possible to explicitly calculate the probabilities of all the configurations on a tree. This also helps to develop an intuition for other graphs. Trees are also useful in application, such as to model evolution in biology.

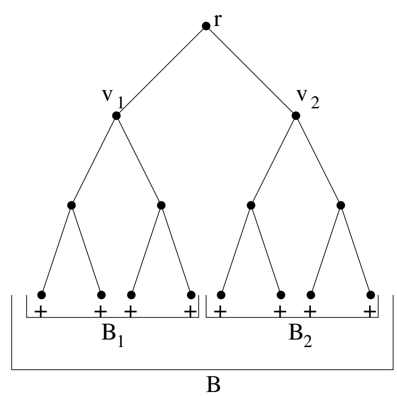

Figure 3: Ising model on a tree

Consider an Ising model distribution on a tree, with all the leaves (set B) fixed to spins +1. Say we now want to compute the probability that the spin σ_r of the root is +1 (We henceforth use + to mean +1 and - for -1). Thus, we want to compute $\mathbf{P}[\sigma_r = +|\sigma_B = +]$. By Bayes' rule

$$\frac{\mathbf{P}[\sigma_r = +|\sigma_B = +]}{\mathbf{P}[\sigma_r = -|\sigma_B = +]} = \frac{\mathbf{P}[\sigma_B = +|\sigma_r = +] \cdot \mathbf{P}[\sigma_r = +]}{\mathbf{P}[\sigma_B = +|\sigma_r = -] \cdot \mathbf{P}[\sigma_r = -]} = \frac{\mathbf{P}[\sigma_B = +|\sigma_r = +]}{\mathbf{P}[\sigma_B = +|\sigma_r = -]}$$

where the last equality follows by symmetry. Hence it suffices to calculate $\mathbf{P}[\sigma_B = +|\sigma_r = +]$ and $\mathbf{P}[\sigma_B = +|\sigma_r = -]$. Also,

$$\begin{aligned}
\mathbf{P}[\sigma_B = +|\sigma_r = +] &= \sum_{b_1,b_2 \in \{-,+\}} \mathbf{P}[\sigma_B = +, \sigma_{v_1} = b_1, \sigma_{v_2} = b_2|\sigma_r = +] \\
&= \sum_{b_1,b_2 \in \{-,+\}} \mathbf{P}[\sigma_{B_1} = +, \sigma_{v_1} = b_1|\sigma_r = +] \cdot \mathbf{P}[\sigma_{B_2} = +, \sigma_{v_2} = b_2|\sigma_r = +]
\end{aligned}$$

Finally

$$\begin{aligned}\mathbf{P}[\sigma_{B_1} = +, \sigma_{v_1} = b_1 | \sigma_r = +] &= \mathbf{P}[\sigma_{v_1} = b_1 | \sigma_r = +] \cdot \mathbf{P}[\sigma_{B_1} = + | \sigma_{v_1} = b_1, \sigma_r = +] \\ &= \mathbf{P}[\sigma_{v_1} = b_1 | \sigma_r = +] \cdot \mathbf{P}[\sigma_{B_1} = + | \sigma_{v_1} = b_1]\end{aligned}$$

The last equality follows from the fact that anything in the subtree rooted at v_1 is independent of r after conditioning on v_1. Now $\mathbf{P}[\sigma_{v_1} = b_1 | \sigma_r = +]$ can be directly computed and $\mathbf{P}[\sigma_{B_1} = + | \sigma_{v_1} = b_1]$ is an expression of same type as we started with, only on a smaller tree. Hence, one can compute the required probability by recursing down the tree.