Lecture Notes — April 26, 2017

Julien Clancy

May 1, 2017

In this lecture we begin analyzing the phase transition in the Ising model. Specifically, the Ising model is: we have a graph $G = (V, E)$ where $|V| = n$ and a function $x\colon V \to \{-1, +1\}$ — that is, a vector $x \in \{-1, +1\}^n$. We sample x according to the distribution

$$p(x) = \frac{1}{Z} \exp\left[\gamma \sum_{(i,j)\in E} x_i x_j + h \sum_i x_i\right]$$

We think of $\gamma > 0$, so there is a higher probability of $x_i = x_j$ when $i \sim j$ than of the opposite. Here as usual Z is a normalizing constant. The other model we consider, the Curie-Weiss model, is a special case where $G = K_n$ (and we reparameterize γ):

$$p(x) = \frac{1}{Z} \exp\left[\frac{\beta}{n} \sum_{i\neq j} x_i x_j + h \sum_i x_i\right]$$

We take $\beta > 0$. If $h \geq 0$ then the event $x_i = 1$ for all i is the most probable. This model is useful since we can reparameterize it in terms of its "magnetization" $\overline{x} = \frac{1}{n}\sum x_i$ so that the parameter in the exponent is

$$\frac{n\beta}{2}\overline{x}^2 - \frac{\beta}{2} + nh\overline{x}$$

Then the probability of a configuration depends only on its magnetization. We now have three questions before us, in this simple model:

1. What is the typical value of $\overline{x}$? Lest this question seem trivial — because the highest-probability even is all ones — consider that we might typically be one of a class of many indistinguishable configurations, as the high-dimensional normal distributions are typically on a spherical shell even as the origin has the highest probability mass. It will turn out here that if $\beta < \beta_c = 1$, then typically $\overline{x} \approx 0$, while if $\beta > \beta_c = 1$ then $\overline{x}$ is the solution to $\overline{x} = \tanh \beta\overline{x}$.

2. What is the mixing time of the Glauber dynamics? It will turn out that in the subcritical regime ($\beta < \beta_c$) it takes about $n \log n$ time to mix, while in the supercritical regime the mixing time is no less than $e^{\Omega(n)}$.

3. What is the overall structure of the probability measure, geometrically speaking? In an unsurprising way given the last assertion, in the subcritical regime the measure gives probability mass to a unified blob, while in the supercritical regime there are at least two almost-disconnected components that receive substantial mass.

Let's briefly state what the Glauber dynamics are here: we choose a vertex $v \in V$ randomly, then update its spin to be $+1$ with probability $\mathbb{P}[x_v = 1 \mid x_{\sim v}]$, and -1 otherwise. This former probability is

$$\frac{\exp\left[\gamma \sum_{v\sim w} x_w + h\right]}{\exp\left[\gamma \sum_{v\sim w} x_w + h\right] + \exp\left[-\gamma \sum_{v\sim w} x_w + h\right]}$$

First we tackle the question of what the typical value of $\overline{x}$ is.

Lemma 1. *Let $m \in [-1, 1]$, and define*

$$\psi_\beta(m) = \frac{\beta}{2}m^2 + hm + H\left(\frac{m+1}{2}\right)$$

where $H(x) = -x\log x - (1-x)\log(1-x)$ is the binary entropy. Then

$$\frac{e^{-\beta/2}}{m+1}\frac{e^{n\psi_\beta(m)}}{Z_n(\beta)} \leq \mathbb{P}[\overline{x} = m] \leq \frac{e^{n\psi_\beta(m)}}{Z_n(\beta)}$$

Proof. By simple counting

$$\mathbb{P}[\overline{x} = m] = \frac{1}{Z_n(\beta)} \exp\left[n\left(\frac{\beta}{2}m^2 - \frac{\beta}{2n} + hm\right)\right] \cdot \binom{n}{n(m+1)/2}$$

By Stirling's approximation from both sides

$$\binom{n}{n(m+1)/2} \sim \exp\left[nH\left(\frac{m+1}{2}\right)\right]$$

where on one side we have a simple inequality and on the other we have it with the factor of $\frac{1}{m+1}$. Inserting this concludes. □

In light of the previous lemma, the following has the interpretation that the events with the highest probability make overwhelming contributions to the average case, while the other events make almost none. This is just how in the limit

$$\left(\sum e^{nb_i}\right)^{1/n} \to e^{\max_i b_i}$$

Lemma 2. *Define the "free energy"*

$$\varphi_n(\beta) = \frac{1}{n} \log Z_n(\beta)$$

and

$$\varphi_*(\beta) = \sup_{m \in [-1,1]} \psi_\beta(m)$$

Then

$$\varphi_*(\beta) - \frac{\beta}{2n} - \frac{1}{n} \log n(n+1) \leq \varphi_n(\beta) \leq \varphi_*(\beta) + \frac{1}{n} \log n$$

Proof. By the last lemma we have

$$Z_n(\beta) \leq \sum_m e^{n\psi_\beta(m)} \leq n \sum_m e^{n\varphi_*(\beta)}$$

and we can use the above scaling limit to conclude, for very large n. The lower bound is the same. To get the result for all n, not just large ones, we just muck around with the limit. □

Now we are in a position to answer the first question we posed at the beginning of the lecture, by maximizing $\psi_\beta(m)$. We have

$$\psi'_\beta(m) = h + \beta m + \log\left(\frac{1-m}{1+m}\right) = 0$$
$$\Rightarrow m = \tanh(\beta m + h)$$

The behavior of this solution differs dramatically for $\beta \leq 1$ and $\beta > 1$. For $\beta \leq 1$ there is only one solution, corresponding to a unique maximum of ψ_β (it is off-center). For $\beta > 1$ (and $0 \leq h \leq h^*(\beta)$) there are actually three solutions, because the slope of the tanh function becomes steeper than the line $y = m$, and these correspond to two maxima and one minimum. The minimum is at zero, and the maxima are symmetric about the origin, as illustrated in the next theorem. The larger of the two maxima is on the right, and we call it $m^+(\beta, h)$.

If $\beta > 1$ but $h > H^*(\beta)$ there is agaon only one local maximum; we call this m^* still.

Theorem 1. *If $h > 0$ or $h = 0$ and $\beta \leq 1$, there is $C = C_\varepsilon$ such that*

$$\mathbb{P}[|\overline{x} - m^*(\beta, h)| < \varepsilon] \geq 1 - e^{-C_\varepsilon n}$$

If $h = 0$ and $\beta > 1$, then

$$\mathbb{P}[|\overline{x} - m^*(\beta, h)| < \varepsilon] > 1/2 - e^{-C_\varepsilon n}$$

and

$$\mathbb{P}[|\overline{x} + m^*(\beta, h)| < \varepsilon] > 1/2 - e^{-C_\varepsilon n}$$

Proof. If $h > 0$ or $\beta \leq 1$, then

$$\begin{aligned}\mathbb{P}[|\overline{x} - m^*(\beta, h)| > \varepsilon] &\leq \frac{n+1}{Z_n(\beta)} \exp\left[n \max\{\psi_\beta \mid |m^*_m| > \varepsilon\}\right] \\ &\leq \frac{(n+1)^3}{e^{\beta/2}} \exp\left[n \max\{\psi_\beta - \varphi_*(\beta) \mid |m^*_m| > \varepsilon\}\right]\end{aligned}$$

The term in the exponent is at most $-C_\varepsilon$, and this concludes the first part. The other is similar, and uses the symmetry of ψ_β. $\square$

This answers the first question. We leave the second until later. For the third:
Theorem 2. *Define $\Omega_+ = \{x \mid \overline{x} \geq 0\}$ and $\Omega_- = \{x \mid \overline{x} < 0$, and define the ε-boundary $\partial_\varepsilon(A) = \{x \mid 1 \leq d_H(x, A) \leq n\varepsilon\}$. Then there is $\varepsilon > 0$ such that*

$$\frac{p(\partial_\varepsilon \Omega_\pm)}{p(\Omega_\pm)(1 - p(\Omega_\pm))} \to 0$$

and in fact this occurs exponentially quickly.

Proof. Take $\varepsilon = m^*/2$; then $\partial_\varepsilon(\Omega_+) = \{x \mid -m^*/2 \leq \overline{x} \leq 0\} \Rightarrow p(\partial_\varepsilon \Omega_+) \leq e^{-cn}$ from the last theorem. $\square$

From this, we get a conductance lower bound on the mixing time for Glauber dynamics of $e^{\Omega(n)}$ in the $\beta > 1$ case.

The last thing to do is fast mixing at high temperatures. For this we move to more general graphs, and assume $h = 0$. We'll make use of path coupling:
Theorem 3. *Suppose for all neighboring states $x, y \in \Omega$ there is a coupling $X_1 \sim p(x, \cdot)$, $Y_1 \sim p(y, \cdot)$ such that $\mathbb{E}\rho(X_1, Y_1) \leq e^{-\alpha}\rho(x, y)$, where ρ is a metric on the state space Ω. Then $t_{mix} \lesssim \alpha^{-1} \log diam\Omega$.*

Let σ and τ be two states differing only at some vertex v, and let ρ be the Hamming metric (so $\rho(x, y) = 1$). The coupling we use is obvious: let

$$\begin{aligned}p(\sigma, w) &= \mathbb{P}[\sigma_w = +1 \mid \sigma_{\sim w}] \\ p(\tau, w) &= \mathbb{P}[\tau_w = +1 \mid \tau_{\sim w}]\end{aligned}$$

and let $U \in [0, 1]$. We pick w randomly and update $\tau(w)$ as 1 if $U \leq p(\tau, w)$ and -1 else, and update $\sigma(w)$ as 1 if $U \leq p(\sigma, w)$ and -1 else. After one step, if we selected w that is not v or a neighbor of v then $\rho(\sigma', \tau')$ stays the same, at 1. If we selected $w = v$ then $\rho(\sigma', \rho') = 0$, and if we selected $w \sim v$ then with probability $p(\tau, w) - p(\sigma, w)$ we have $\rho(\sigma', \tau') = 2$ and with the complementary probabiltiy the distance remains 1. Then

$$\mathbb{E}\rho(\sigma', \tau') = 1 - \frac{1}{n} + \frac{1}{n} \sum_{w \sim v} p(\tau, w) - p(\sigma, w)$$

If we let $S(w) = \sum_{u \sim w} \sigma(u)$ then $S + 2 = \sum_{u \sim w} \tau(u)$ and so

$$p(\tau, w) - p(\sigma, w) = \frac{e^{\beta(S+2)}}{e^{\beta(S+2)} + e^{-\beta(S+2)}} - \frac{e^{\beta S}}{e^{\beta S} + e^{-\beta S}} = \frac{1}{2} (\tanh \beta(S+2) - \tanh \beta S) \leq \tanh \beta$$

so

$$\mathbb{E}\rho(\sigma', \tau') \leq 1 - \frac{1}{n} + \frac{\Delta}{n} \tanh \beta = 1 - C(\beta) \leq e^{-C(\beta)/n}$$

so we can apply the path coupling theorem with $\alpha = C(\beta)/n$ to get a mixing time of $n/C(\beta) \log n$, since the diameter of the state space is clearly $\log n$. Here we defined $C(\beta) - 1 = \Delta \tanh \beta$, where Δ is the maximum degree.