# Markov Chains and Algorithmic Applications: WEEK 13

## 1 Ising model

This week, we review this famous model from statistical physics, which has also found numerous applications outside of physics, in particular in computer science.

### 1.1 Model description

Let us first describe the general model. Consider an undirected graph $G = (V, E)$, with $|V| = N$ vertices, as already seen for the coloring problem. On each vertex $v \in V$, define a variable $x_v$ taking values in $\{+1, -1\}$, representing in physics the small magnetic moment of an atom, or "spin", which can take values "up" $(+1)$ or "down" $(-1)$. The vector $x = (x_v,\, v \in V)$ represents the state of the system, which belongs to the state space $S = \{+1, -1\}^V$. On this state space $S$, define the following energy function (a.k.a. *Hamiltonian* function)

$$H(x) = -\sum_{(v,w)\in E} J_{vw}\, x_v\, x_w - \sum_{v\in V} h_v\, x_v$$

where $J_{vw} = J_{wv} \in \mathbb{R}$ and $h_v \in \mathbb{R}$; $J_{vw}$ represents the interaction between spins $x_v$ and $x_w$, and $h_v$ represents the strength of the external magnetic field at vertex $v$.

Physical systems naturally tend to minimize energy, so we should be looking for configurations $x \in S$ minimizing the above function. But it is also natural to assume a non-zero temperature, which creates fluctuations in the system and prevents it from always staying in the global minima of the energy function $H(x)$. A reasonable model is to assume that, at temperature $T > 0$, the distribution of the system is given by the *Gibbs distribution*:

$$\pi_\beta(x) = \frac{\exp(-\beta H(x))}{Z_\beta}, \quad x \in S$$

where $\beta = 1/T > 0$ is the inverse of the temperature $T$ and $Z_\beta = \sum_{x\in S} \exp(-\beta H(x))$ is the corresponding normalization constant (a.k.a. in physics as the *partition function*).

Characterizing the properties of the above distribution is a key question in physics[1], where the number of atoms considered is usually of the order of the Avogadro number ($N \simeq 6 \cdot 10^{23}$), so considering the limit $N \to \infty$ certainly makes sense in this context ! What is of particular interest to physicists is the quantity (a.k.a. *observable*) called the *magnetization* of the system, given by

$$M(x) = \frac{1}{N}\sum_{v\in V} x_v$$

Note that by assumption $M(x) \in [-1, +1]$. In particular, one is interested in the distribution of this magnetization under the distribution $\pi_\beta$, as well as in its average (at inverse temperature $\beta$), denoted as

$$\langle M \rangle_\beta = \sum_{x\in S} M(x)\, \pi_\beta(x)$$

which is a macroscopic quantity that can be measured experimentally and compared with the predictions of the various models.

Let us now review two "simple" models (but as you will notice, simple models can lead to the description of very interesting phenonema, as it is often the case in probability).

[1] But please note that computing the partition function $Z_\beta$ alone is already a challenge in general...

### 1.1.1 The Curie-Weiss model

In this model, it is assumed that the graph $G$ is complete and that $J_{vw} \equiv J/N > 0$ for all $(v, w) \in E$ and $h_v \equiv h$ for all $v \in V$. Assuming that the coefficients $J_{vw}$ are positive is referred to in physicis as a *ferromagnetic model*: in this case, spins $x_v$ and $x_w$ tend to align (i.e., to take equal values, either $+1$ or $-1$, in order to minimize the energy) [2].

It turns out in this case that the Hamiltonian $H$ can be expressed as a simple function of the magnetization $M$, namely

$$H(x) = -N\left(\frac{JM(x)^2}{2} + hM(x)\right) + \frac{J}{2}$$

which leads to a relatively easy analysis of the distribution of the magnetization $M$ (at inverse temperature $\beta$). The conclusions are as follows:

1. When $h = 0$ (no external magnetic field), it is the case at all temperatures that the average magnetization $\langle M \rangle_\beta = 0$ (which is somewhat to be expected...), but the distribution of this magnetization depends highly on the temperature:

1a. If $\beta J < 1$ (*high temperature regime*), then the distribution of $M$ is a bell-shaped curve centered in 0, as illustrated below on Figure 1 (left). Please note that the peak gets narrower as $N$ increases. In this regime, the high temperature wins over the interactions between the spins, so no global alignment is found.

1b. If $\beta J > 1$ (*low temperature regime*), then the distribution of $M$ becomes bimodal, as illustrated again on Figure 1 (right), and the peaks get again narrower as $N$ increases. In this regime, the ferromagnetic interaction between the spins wins over the temperature, so a majority of spins tend to align in one direction, but as there is no external magnetic field, the chosen direction for the alignement is randomly in the plus or minus direction., Overall, the average of the magnetization remains therefore zero.

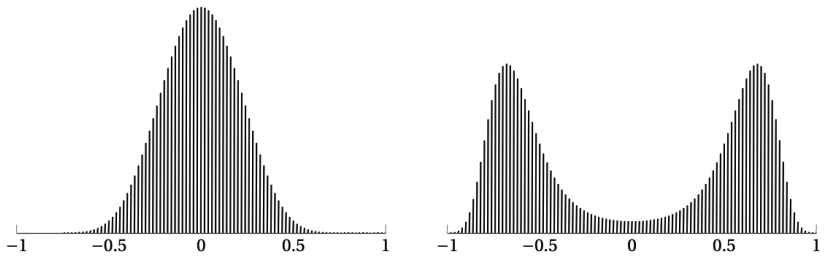


Figure 1: Distribution of the magnetization $M$ in the absence of external magnetic field ($h = 0$), for $N = 100$ spins, when $\beta J = 0.8$ (left) and $\beta J = 1.2$ (right). Figure taken from S. Friedli and Y. Velenik, *Statistical Mechanics of Lattice Systems: a Concrete Mathematical Introduction*, Cambridge University Press, 2017, `https://www.unige.ch/math/folks/velenik/smbook`.

What is observed here is called a *phase transition* in physics; a completely different different behaviour of the system is observed when some parameter (here $\beta J$) is below or above a given value (1).

2. When $h > 0$ (please note that the situation is symmetric when $h < 0$), the situation changes in the following sense. First, the average magnetization $\langle M \rangle_\beta > 0$, and the above distributions become now tilted in the direction of the external magnetic field:

2a. When $\beta J < 1$ (*high temperature regime*), the bell-shaped curve gets tilted towards the right, with a maximum at some value $m > 0$.

[2] The factor $1/N$ is assumed here as a normalization factor, to compensate for the fact that the graph is complete and that each spin interacts with order $N$ nodes.

2b. When $\beta J > 1$ (*low temperature regime*), the left-hand side peak gets lower and the right-hand side peak gets higher, so that now the majority of spins tend to privilege the plus direction.

As a final remark, note that even if the above model is somewhat unrealistic (the fact that interaction takes place between *all* pairs of spins is certainly unrealistic), it leads to a description of a phenomenon that is observed in practice ! This category of models, where interactions between spins are somehow averaged out across all pairs, are called *mean-field models*. They are interesting, because amenable to analysis and at the same time surprisingly rich to describe natural phenomena.

### 1.1.2 The Ising-Lenz model

The original model proposed in 1920 by Wilhelm Lenz to his student Ernst Ising was the model where the graph $G$ is the $d$-dimensional grid (i.e., $V = \mathbb{Z}^d$ and the only existing edges are the ones between nearest neighbours) and it is assumed that $J_{vw} \equiv J \in \mathbb{R}$ for all $(v, w) \in E$ and $h_v \equiv h$ for all $v \in V$. Ising studied the one-dimensional case and concluded that no phase transition occurs, even in the limit $N \to \infty$ ! For some reason, only the name of Ising remained attached to the model...

It turns out that the two-dimensional model is much harder to study, and it is only in 1944 that Lars Onsager found an analytic description of the model, showing that a phase transition occurs in this case in the large $N$ limit. The three-dimensional model is even harder, so less is known about this model, but a phase transition has also been proven to occur in this case also. Surprisingly perhaps, things become again easier in dimension four and above (but the pertinence of the model to physics diminishes a little...).

## 1.2 MCMC sampling

The Metropolis algorithm is of course well suited to sample from the distribution $\pi_\beta$ in the general case. Start from a uniformly random state $X_0 \in S = \{+1, -1\}^V$ and consider the base chain with transition probabilities

$$\psi_{xy} = \begin{cases} 1/N & \text{if } x \sim y \text{ (i.e., } x \text{ and } y \text{ differ in exactly one vertex } u \in V) \\ 0 & \text{otherwise} \end{cases}$$

This chain is clearly irreducible (and also periodic with period 2, but this is in general not an issue: see below) and the matrix $\psi$ is symmetric. So the conditions for the Metropolis algorithm to work are met, and the acceptance probabilities are given by

$$a_{xy} = \min\{1, \pi_\beta(y)/\pi_\beta(x)\} = \min\{1, \exp(-\beta(H(y) - H(x))\}$$

As $y \sim x$ (and say the only discrepancy takes place at vertex $u \in V$, where $y_u = -x_u$), we obtain

$$\begin{aligned} H(y) - H(x) &= -\sum_{(v,w)\in E} J_{vw}\, y_v\, y_w - \sum_{v\in V} h_v\, y_v + \sum_{(v,w)\in E} J_{vw}\, x_v\, x_w + \sum_{v\in V} h_v\, x_v \\ &= 2x_u \sum_{w\in V\,:\,(u,w)\in E} J_{uw}\, x_v + 2h_u\, x_u = 2x_u \left( \sum_{w\in V\,:\,(u,w)\in E} J_{uw}\, x_v + h_u \right) = 2x_u\, h_u^{(loc)} \end{aligned}$$

where $h_u^{(loc)} = \sum_{w\in V\,:\,(u,w)\in E} J_{uw}\, x_w + h_u$ denotes the *local magnetic field* at vertex $u$. This implies that

$$a_{xy} = \begin{cases} \exp(-2\beta x_u\, h_u^{(loc)}) & \text{if } x_u\, h_u^{(loc)} > 0 \\ 1 & \text{if } x_u\, h_u^{(loc)} \le 0 \end{cases}$$

This says naturally that if the value of the spin $x_u$ is aligned with that of the local magnetic field $h_u^{(loc)}$, then the proposed move $x \to y$ will actually increase the energy $H$, so one should make this move with

some positive probability only; while if $x_u$ and $h_u^{(loc)}$ are opposed, then one should always accept the proposed move, which will decrease the energy $H$.

From there, the transition probabilities $p_{xy}$ of the Metropolis chain can be easily computed, and Theorem 2.2 seen in Week 9 shows that $\pi_\beta$ is the stationary distribution of this chain[3]. Please note also that here, we should not choose $\beta$, are this parameter is given by the physics of the problem, more precisely by the inverse of the temperature of the considered system.

## 1.3 Gibbs sampling

Gibbs sampling (a.k.a. *Glauber dynamics*) is a variant of the Metropolis algorithm. The idea is the following: starting from a state $x \in S$, choose a vertex $u \in V$ uniformly at random and define two states $x^{(u,\pm)}$ by

$$x_u^{(u,\pm)} = \pm 1 \quad \text{and} \quad x_v^{(u,\pm)} = x_v \quad \forall v \neq u$$

The chain then makes a move towards state $x^{(u,\pm)}$ with probability $p_{x,x^{(u,\pm)}} = \frac{1}{N} \frac{\pi_\beta(x^{(u,\pm)})}{\pi_\beta(x^{(u,+)})+\pi_\beta(x^{(u,-)})}$ (where the factor $1/N$ accounts for the random choice of the vertex $u$). Observe that these transition probabilities are independent of the current state $x_u$ of vertex $u$; in particular, there is always a positive probability that the state of the chain does not change, similarly to the Metropolis algorithm.

It is quite straightforward to check (under the same set of assumptions as for the Metropolis algorithm) that such a chain converges towards the limiting and stationary distribution $\pi_\beta$, as the detailed balance equation is satisfied, namely

$$\pi_\beta(x)\, p_{xy} = \pi_\beta(y)\, p_{yx} \quad \forall x, y \in S$$

**Claim.** The computation of the above probabilities $p_{x,x^{(u,\pm)}}$ gives

$$p_{x,x^{(u,\pm)}} = \frac{1}{2N}\left(1 \pm \tanh(\beta h_u^{(loc)})\right) \quad \text{where recall that} \quad h_u^{(loc)} = \sum_{w \in V\,:\,(u,w)\in E} J_{uw}\, x_w + h_u$$

Let us prove the claim for $p_{x,x^{(u,+)}}$ (the proof for $p_{x,x^{(u,-)}}$ being symmetric):

$$p_{x,x^{(u,+)}} = \frac{1}{N} \frac{\pi_\beta(x^{(u,+)})}{\pi_\beta(x^{(u,+)}) + \pi_\beta(x^{(u,-)})} = \frac{\exp(-\beta H(x^{(u,+)}))}{\exp(-\beta H(x^{(u,+)})) + \exp(-\beta H(x^{(u,-)}))}$$

Note that here also, the normalization constant $Z_\beta$ is disappearing, which ensures the feasible implementation of the algorithm in reasonable time. Next, we obtain

$$p_{x,x^{(u,+)}} = \frac{1}{N} \frac{1}{1 + \exp(-\beta(H(x^{(u,-)}) - H(x^{(u,+)})))}$$

but we computed earlier (cf. previous section) that

$$H(x^{(u,-)}) - H(x^{(u,+)}) = +2h_u^{(loc)}$$

so after some algebra

$$p_{x,x^{(u,+)}} = \frac{1}{N} \frac{1}{1 + \exp(-2\beta h_u^{(loc)})} = \frac{1}{N} \frac{\exp(+\beta h_u^{(loc)})}{\exp(+\beta h_u^{(loc)}) + \exp(-\beta h_u^{(loc)})} = \frac{1}{2N}\left(1 + \tanh(\beta h_u^{(loc)})\right)$$

which proves the claim. □

[3] As already seen before, the periodicity of the base chain is not a problem, as soon as some acceptance probabilities $a_{xy}$ are strictly below 1, which happens for any non-constant energy function $H(x)$.

## 1.4 Exact simulation

In the following, we will see that for the Ising ferromagnetic model on an arbitrary graph $G$, exact simulation is doable in reasonable time.

Consider the Gibbs sampling algorithm seen in the previous section. We will use the following random mapping representation for it:

$$X_{n+1} = \Phi(X_n, U_{n+1}^{(1)}, U_{n+1}^{(2)}) \quad \text{where} \quad \Phi(x, u_1, u_2)_v = \begin{cases} +1 & \text{if } v = u_1 \text{ and } 0 \le u_2 \le \frac{1}{2}\left(1 + \tanh(\beta h_{u_1}^{(loc)})\right) \\ -1 & \text{if } v = u_1 \text{ and } \frac{1}{2}\left(1 + \tanh(\beta h_{u_1}^{(loc)})\right) < u_2 \le 1 \\ x_v & \text{if } v \neq u_1 \end{cases}$$

Note that we need here two sequences of i.i.d. random variables $(U_n^{(1)},\, n \ge 1)$ and $(U_n^{(2)},\, n \ge 1)$, the $U^{(1)}$'s being uniformly distributed on $V$ and the $U^{(2)}$'s being uniformly distributed on $[0,1]$.

Next, consider the following partial order on $S = \{+1, -1\}^V$:

$$x \preceq y \quad \Leftrightarrow \quad x_v \le y_v, \quad \forall v \in V$$

In particular, let us denote by $\underline{x} = (-1, \ldots, -1)$ the smallest of all configurations and by $\overline{x} = (+1, \ldots, +1)$ the largest of all configurations (with respect to this partial order). Clearly, $\underline{x} \preceq x \preceq \overline{x}$ for all $x \in S$.

**Claim.** On an undirected graph $G = (V, E)$, if all the interactions are ferromagnetic (i.e., $J_{vw} \ge 0$ for all $(v, w) \in E$), then the above random mapping representation of the Gibbs sampler is monotone, i.e.:

$$x \preceq y \quad \Rightarrow \quad \Phi(x, u_1, u_2) \preceq \Phi(y, u_1, u_2) \quad \forall x, y \in S,\ u_1 \in V,\ u_2 \in \mathbb{R}$$

Then in the Propp-Wilson algorithm, all trajectories are sandwiched between the two trajectories emanating from $\underline{x}$ and $\overline{x}$. Therefore, to check coalescence, it is enough to check for coalescence for these two trajectories only (instead of the $2^N$ trajectories corresponding to all spin configurations), which is doable in reasonable time.

*Proof of the claim.* Let us consider $x \preceq y$ and $u_1 \in V$, $u_2 \in [0,1]$ and choose also $v \in V$.

By the above definition of $\Phi$, if $v \neq u_1$, then because $x_v \le y_v$ by assumption, we also have

$$\Phi(x, u_1, u_2)_v = x_v \le \Phi(y, u_1, u_2)_v = y_v$$

Consider now the case $v = u_1$. The only problem that could occur in this case is that $\Phi(x, u_1, u_2)_v = +1$ and $\Phi(y, u_1, u_2)_v = -1$, which would not preserve the partial order $\preceq$. But note that

$$\Phi(x, u_1, u_2)_v = +1 \quad \text{iff} \quad 0 \le u_2 \le \frac{1}{2}\left(1 + \tanh(\beta h_{u_1}^{(loc,x)})\right)$$

where $h_{u_1}^{(loc,x)} = \sum_{w \in V\,:\,(u_1,w) \in E} J_{u_1,w}\, x_w + h_{u_1}$, and

$$\Phi(y, u_1, u_2)_v = -1 \quad \text{iff} \quad \frac{1}{2}\left(1 + \tanh(\beta h_{u_1}^{(loc,y)})\right) < u_2 \le 1$$

where $h_{u_1}^{(loc,y)} = \sum_{w \in V\,:\,(u_1,w) \in E} J_{u_1,w}\, y_w + h_{u_1}$

The fact is that, because the interactions are ferromagnetic and $x \preceq y$ by assumption, we have

$$h_{u_1}^{(loc,x)} - h_{u_1}^{(loc,y)} = \sum_{v \in V\,:\,(u_1,w) \in E} J_{u_1,w}\,(x_w - y_w) \le 0$$

so it is never the case that $u_2 \le \frac{1}{2}\left(1 + \tanh(\beta h_{u_1}^{(loc,x)})\right)$ and $u_2 > \frac{1}{2}\left(1 + \tanh(\beta h_{u_1}^{(loc,y)})\right)$ at the same time. This proves the claim. □