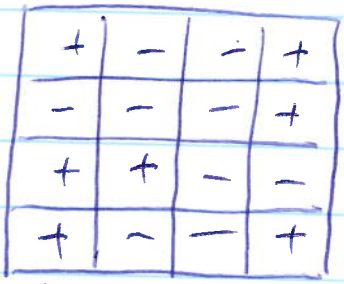
Application in Statistical Mechanics (or: Physics)

Ising Model and Markov Chain Monte Carlo for Ising Model

German physicist Ernst Ising (1900–1998).
Magnetization – statistical properties.

L = # sites

$S = \{\sigma = (\sigma_1, \dots, \sigma_L) \in \mathbb{R}^L : \sigma_i = 1 \text{ or } -1,\ 1 \le i \le L\}$



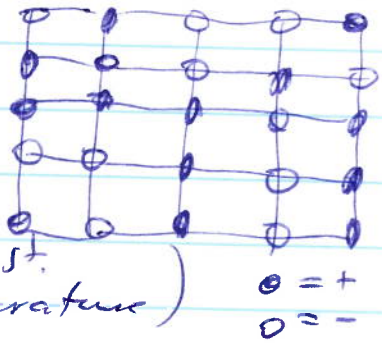
Hamiltonian

$$H(\sigma) = -\sum_{(i,j)} \sigma_i \sigma_j = -\sum_{(i,j),} \sigma_i \sigma_j \quad (J = 1)$$

(i,j) denotes nearest-neighbor pairs

[More general $H(\sigma) = -J \sum_{(i,j)} \sigma_i \sigma_j - h \sum \sigma_i$
The h-term describes the external field.
J: coupling const.]

or use (better)



Target distribution

$$\pi_\beta(\sigma) = \frac{1}{Z_\beta} e^{-\beta H(\sigma)}$$

$\beta = \frac{1}{k_B T}$ (k_B = Boltzmann const.
T = absolute temperature)

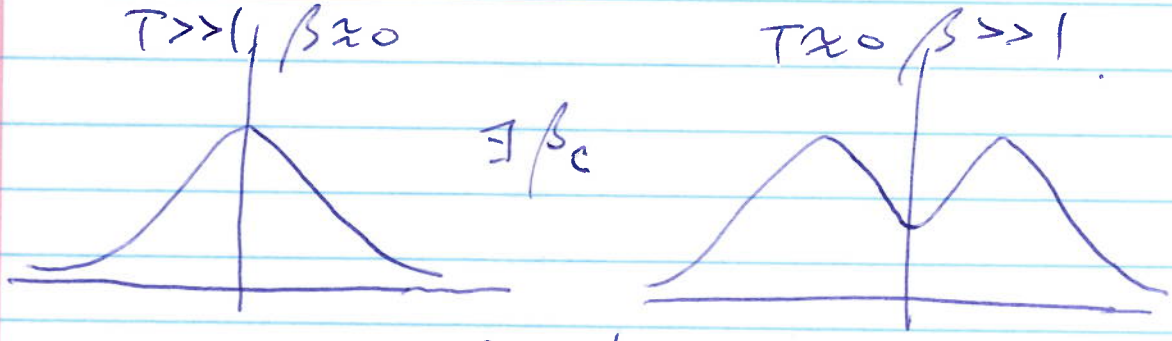
$$Z_\beta = \sum_{\sigma \in S} e^{-\beta H(\sigma)}$$: a normalizing constant, called the partition function.

Problems of interest.

- Distribution of spins 1 and spins -1. (up or down)
- Interface defined by +1, -1
- $T \to 0$ or $T \to \infty$: limiting behavior.
- $L \to \infty$. limiting behavior.
- ...

Phase transition for the mean

$$M(\sigma) = \frac{1}{L} \sum_{i=1}^{L} \sigma_i$$



distribution of mean

Usually, Z_β is hard to calculate.

The Metropolis MCMC Algorithm

General a Markov chain $X_0, X_1, \ldots, X_n \ldots$ with all $X_k \in S$.

Then $\pi^\beta(\sigma) = \lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} \{X_k = \sigma\} \; \mathbb{I}\{X_k = \sigma\}$

Given $X_k = \sigma \in S$. $\sigma = (\sigma_1, \cdots, \sigma_L)$.

① Pick $J \in \{1, \cdots, L\}$ uniformly at random.

② Flip. Set

$$Y \leftarrow (\sigma_1, \cdots \sigma_{J-1}, -\sigma_J, \sigma_{J+1}, \cdots, \sigma_L)$$

③ Set $\alpha = \min\left(1, \frac{\pi^\beta(Y)}{\pi^\beta(X_k)}\right)$

$$= \min\left(1, \frac{e^{-\beta H(Y)}}{e^{-\beta H(X_k)}}\right)$$

④ Accept Y (i.e., $X_{k+1} \leftarrow Y$) with the probability α.

generate $U \sim \mathcal{U}[0,1]$

If $U \le \alpha$ then $X_{k+1} \leftarrow Y$;

If $U > \alpha$ then $X_{k+1} \leftarrow X_k$.

Note: $\frac{e^{-\beta H(Y)}}{e^{-\beta H(X_k)}} = e^{-\beta[H(Y) - H(X_k)]}$

This only involves the nearest neighbors of J.

So, the calculation of $\Delta H = H(Y) - H(X_k)$ is local.

The Gibbs Sampler

Given $X_k = \sigma = (\sigma_1, \cdots, \sigma_L) \in S$

① Pick $J \in \{1, \cdots, L\}$ uniform at random.

(2) Define

$$\sigma[+] = (\sigma_1, \cdots, \sigma_{j-1}, 1, \sigma_{j+1}, \cdots, \sigma_L) \in S$$

$$\sigma[-] = (\sigma_1, \cdots, \sigma_{j-1}, -1, \sigma_{j+1}, \cdots, \sigma_L) \in S$$

$$p_k(\sigma, \sigma[+]) = \frac{e^{-\beta H(\sigma[+])}}{e^{-\beta H(\sigma[+])} + e^{-\beta H(\sigma[-])}}$$

$$p_k(\sigma, \sigma[-]) = \frac{e^{-\beta H(\sigma[-])}}{e^{-\beta H(\sigma[+])} + e^{-\beta H(\sigma[-])}}$$

(3) Accept $\sigma[+]$ with probability $p_k(\sigma, \sigma[+])$, equivalently, accept $\sigma[-]$ with probability $p_k(\sigma, \sigma[-])$.

General $U \sim U[0,1]$.

If $U \leq p_k(\sigma, \sigma[+])$, $X_{k+1} \leftarrow \sigma[+]$.

If $U > p_k(\sigma, \sigma[+])$ $X_{k+1} \leftarrow \sigma[-]$.

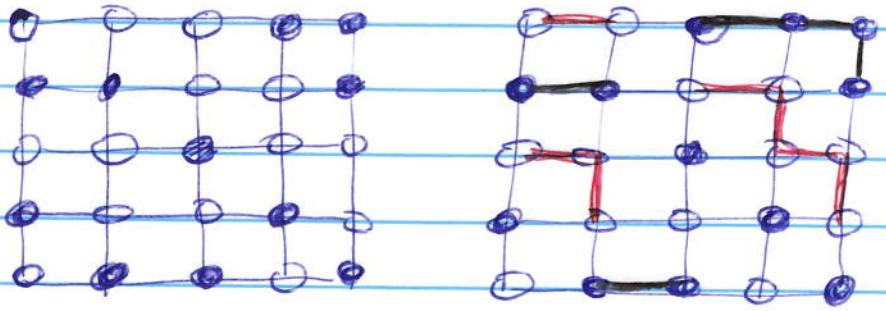
Note: $p_k(\sigma, \sigma[\pm])$ is the transition probability from σ to $\sigma[\pm]$.

Potts model

$$S = \{\sigma = (\sigma_1, \cdots \sigma_L) : \sigma_i \in \{1, \cdots, K\}, 1 \leq i \leq L\}.$$

The Swendsen-Wang Algorithm

This is a clustering algorithm. First, form clusters of sites defined by bonds (auxiliary variables). Second, perturb a cluster of site each time.



● = +
○ = −

clusters connected by red / black edges.

Let $\sigma = (\sigma_1, \cdots, \sigma_L)$ be given.

① Let (i,j) be a pair of neighbors
If $\sigma_i = \sigma_j$ then create a bond $z_{ij} = 1$ with probability $1 - e^{-2\beta J}$ $(J=1)$
If $\sigma_i \neq \sigma_j$, then no bond: $z_{ij} = 0$

② A cluster is a collection of sites connected by bonds. From ①, we form clusters. In each cluster, all sites have the same sign.

③ Assign to each cluster with a random value $+1$ or -1.

Now, let us be a bit more rigorous. Denote by E the set of all neighboring sites:

$$E = \{(i,j) : 1 \le i, j \le L,\ i \ne j,\ \text{sites } i \text{ and } j \text{ are neighbors}\}.$$

Introduce an auxiliary variable

$$z = \{z_{ij} : (i,j) \in E\},$$

where each $z_{ij} = 0$ or 1. i.e., $z \in \{0,1\}^E$. We create a joint distribution $p(\sigma, z)$ such that

(1) The marginal distribution of σ is exactly the Ising distribution, and

(2) It is easy to generate samples from the conditional distributions $p(\sigma|z)$ and $p(z|\sigma)$.

By (2), we can perform the Gibbs sampling of $\sigma|z$ and $z|\sigma$ to obtain a Markov chain

$$X_k = (\sigma^{(k)}, z^{(k)})$$

with limiting distribution $p(\sigma, z)$. Hence, by (1), the limiting distribution of $\sigma^{(k)}$ is the Ising distribution $\pi^\beta(\sigma)$.

To define $p(\sigma, z)$, let us introduce

$$\gamma(z_{ij}, \sigma_i, \sigma_j) = \begin{cases} e^{-\beta} & \text{if } z_{ij} = 0, \\ e^{\beta} - e^{-\beta} & \text{if } z_{ij} = 1 \text{ and } \sigma_i = \sigma_j, \\ 0 & \text{if } z_{ij} = 1 \text{ and } \sigma_i \ne \sigma_j. \end{cases}$$

This is defined for any neighboring pair of sites (i,j). Now, we define

$$P(\sigma, z) = \frac{1}{\Gamma_\beta} \prod_{(i,j)\in E} \gamma(z_{ij}, \sigma_i, \sigma_j),$$

where $\sigma = (\sigma_1, \dots, \sigma_L)$ and $z = \{z_{ij} = (i,j) \in E,\ z_{ij} = 0 \text{ or } 1\}$, and Γ_β is a normalizing const.

Remark. The marginal distribution of z is known as the Fortuin-Kasteleyn or random cluster model.

Proposition The marginal distribution of σ is exactly the Ising distribution

$$\pi^\beta(\sigma) = \frac{1}{Z_\beta} e^{-\beta H(\sigma)} \quad (\sigma \in S).$$

Proof We first observe that for any $(i,j) \in E$,

$$\gamma(0, \sigma_i, \sigma_j) + \gamma(1, \sigma_i, \sigma_j) =$$
$$= e^{-\beta} + \begin{cases} e^{\beta} - e^{-\beta} & \text{if } \sigma_i = \sigma_j \\ 0 & \text{if } \sigma_i \neq \sigma_j \end{cases}$$
$$= \begin{cases} e^{\beta} & \text{if } \sigma_i = \sigma_j \\ e^{-\beta} & \text{if } \sigma_i \neq \sigma_j \end{cases}$$
$$= e^{\beta \sigma_i \sigma_j}$$

Now, let $N = |E|$ and order $(i,j) \in E$ as $(i(1), j(1)), (i(2), j(2)), \dots, (i(N), j(N))$. The marginal distribution, $P(\sigma)$, of the joint distribution $P(\sigma, z)$ on $\sigma \in S$ is

$$P(\sigma) = \sum_{\text{all } z} P(\sigma, z) = \frac{1}{\Gamma_\beta} \sum_{\text{all } z} \gamma \prod_{(i,j)\in E} \gamma(z_{ij}, \sigma_i, \sigma_j)$$

where $z = \{z_{ij} : (i,j) \in E,\ z_{ij} = 0 \text{ or } 1\}$. With our notation, we have

$$P(\sigma) = \frac{1}{\Gamma_\beta} \sum_{z_{i(1),j(1)}=0,1} \sum_{z_{i(2),j(2)}=0,1} \cdots \sum_{z_{i(N),j(N)}=0,1} \prod_{k=1}^{N} \gamma\left(z_{i(k),j(k)}, \sigma_{i(k)}, \sigma_{j(k)}\right)$$

$$= \frac{1}{\Gamma_\beta} \sum_{z_{i(1),j(1)}=0,1} \cdots \sum_{z_{i(N-1),j(N-1)}=0,1} \prod_{k=1}^{N-1} \gamma\left(z_{i(k),j(k)}, \sigma_{i(k)}, \sigma_{j(k)}\right) \sum_{z_{i(N),j(N)}=0,1} \gamma\left(z_{i(N),j(N)}, \sigma_{i(N)}, \sigma_{j(N)}\right)$$

$$= \frac{1}{\Gamma_\beta} \sum_{z_{i(1),j(1)}=0,1} \cdots \sum_{z_{i(N-1),j(N-1)}=0,1} \prod_{k=1}^{N-1} \gamma\left(z_{i(k),j(k)}, \sigma_{i(k)}, \sigma_{j(k)}\right) \cdot \left[\gamma(0, \sigma_{i(N)}, \sigma_{j(N)}) + \gamma(1, \sigma_{i(N)}, \sigma_{j(N)})\right]$$

$$= \frac{1}{\Gamma_\beta} \sum_{z_{i(1),j(1)}=0,1} \cdots \sum_{z_{i(N-1),j(N-1)}=0,1} \prod_{k=1}^{N-1} \gamma\left(z_{i(k),j(k)}, \sigma_{i(k)}, \sigma_{j(k)}\right) \cdot e^{\beta \sigma_{i(N)} \sigma_{j(N)}}$$

$$= \cdots = \frac{1}{\Gamma_\beta} e^{\beta \sigma_{i(1)} \sigma_{j(1)}} e^{\beta \sigma_{i(2)} \sigma_{j(2)}} \cdots e^{\beta \sigma_{i(N)} \sigma_{j(N)}}$$

$$= \frac{1}{\Gamma_\beta} e^{-\beta H(\sigma)}$$

Since $P(\sigma)$ ~~and $\pi(\beta(\sigma)$ are both~~ is a distribution, the normalizing constants ~~are the same~~ is

$$\Gamma_\beta = \sum_{\sigma \in S} e^{-\beta H(\sigma)} = Z_\beta$$

Hence, $P(\sigma) = \frac{1}{Z_\beta} e^{-\beta H(\sigma)} = \pi_\beta(\sigma) \quad \forall \sigma \in S.$ □

We now calculate the conditional distributions

First, consider $z|\sigma$. Note that for a given σ, z_{ij}'s are independent, and

[check this!] $$p(z_{ij}=1|\sigma)=\begin{cases}0 & \text{if } \sigma_i \neq \sigma_j,\\ 1-e^{-2\beta} & \text{if } \sigma_i=\sigma_j.\end{cases}$$

(Cf. the definition of $\gamma(z_{ij},\sigma_i,\sigma_j)$) Thus, it is easy to generate the distribution $p(z|\sigma)$.

Consider now the distribution $p(\sigma|z)$ for a given z. Define

$$S(z)=\{\sigma\in S: p(\sigma,z)>0\}$$
$$=\{\sigma\in S: \sigma_i=\sigma_j \text{ for every nbhr } \langle i,j\rangle \text{ such that } z_{ij}=1\}.$$

Elements (i.e., configurations) in $S(z)$ are decomposed into disjoint clusters, each having the same value (+1 or -1), and each is connected. Let $\eta(z)$ be the total number of clusters. Then, the cardinality

$$|S(z)|=2^{\eta(z)}$$

If $\sigma\in S(z)$, Then, by the definition of the joint distribution $p(\sigma,z)$, we have

$$p(\sigma,z)=\frac{1}{\Gamma_\beta}\left(\prod_{(i,j)\in E,\, z_{ij}=0}\gamma(z_{ij},\sigma_i,\sigma_j)\right)\left(\prod_{(i,j)\in E,\, z_{ij}=1}\gamma(z_{ij},\sigma_i,\sigma_j)\right)$$
$$=\frac{1}{\Gamma_\beta}e^{-\beta\left[|E|-\sum_{(i,j)\in E}z_{ij}\right)}\left(e^{\beta}-e^{-\beta}\right)^{\sum_{(i,j)\in E}z_{ij}}$$

independent of $\sigma\in S(z)$!

If $\sigma \notin S(z)$ then $p(\sigma, z) = 0$. Hence

$$p(\sigma|z) = \begin{cases} 2^{-\gamma(z)} & \text{if } \sigma \in S(z), \\ 0 & \text{if } \sigma \notin S(z). \end{cases}$$

Now, to generate σ given z, with such a conditional distribution, one can first identify the clusters in $S(z)$. Then, for each cluster C choose $\sigma_C \in \{1, -1\}$ uniformly at random (independent for each cluster), and set $\sigma_i = \sigma_C$ in that cluster C.

More properties of the Swendsen-Wang clustering algorithm.

- ergodicity [prove it!]
- Detailed balance

$$\frac{p(A \to B)}{p(B \to A)} = e^{-\beta[H(B) - H(A)]} = \frac{\pi^\beta(A)}{\pi^\beta(B)}.$$

The Wolff single-cluster algorithm

[U. Wolff. PRL, 62:361, 1989]

This is very similar to the Swendsen-Wang algorithm, except it only flips one randomly selected cluster.

Given $\sigma = (\sigma_1, \cdots, \sigma_L) \in S$ (so, $\sigma_i = 1$ or -1)

- choose a random site i.
- Considering sites j that are neighboring to i. If $\sigma_j = \sigma_i$ then join site j to cluster with probability $1 - e^{-2\beta J}$
- Repeat the previous step to generate the cluster.
- When the cluster is constructed, flip its spins.

Same properties / features

- Slightly more effective than the Swendsen-Wang algorithm (the average cluster size is large).
- Recursive.
- Statistical property: detailed balance.

The equilibrium distribution of a classical system of many particles

Let $L>0$ and $\Lambda = (-L, L)^3$. Consider N particles $\vec{r}_1, \cdots, \vec{r}_N$ inside Λ, with the periodical boundary conditions. Given an interaction potential

$$U = U(\vec{r}_1, \cdots, \vec{r}_N)$$

e.g.,

$$U(\vec{r}_1, \cdots, \vec{r}_N) = \sum_{\substack{i,j=1 \\ (i \neq j)}}^{N} U_{ij}(|\vec{r}_i - \vec{r}_j|) + \sum_{i=1}^{N} \Phi(\vec{r}_i)$$

where often

$$U_{ij} = U_0$$

is independent of i, j representing pair (i, j) interaction through their distance and $\Phi(\vec{r}_i)$ represents some external (wall) effect.

The equilibrium distribution is

$$\pi_\beta(\vec{R}) = \frac{1}{Z_\beta} e^{-\beta U(\vec{R})}$$

where $\vec{R} = (\vec{r}_1, \cdots, \vec{r}_N)$, $U(\vec{R}) = U(\vec{r}_1, \cdots, \vec{r}_N)$, $\beta = 1/(k_B T)$, Z_β is the partition function.

$$Z_\beta = \int e^{-\beta U(\vec{R})} d\vec{R}$$

Examples of such systems include.

- Simple liquid system — a Lennard-Jones liquid system, where
$$U_{ij}(r) = 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r}\right)^{12} - \left(\frac{\sigma_{ij}}{r}\right)^{6}\right].$$

- A charged system, where the interaction $U_{ij}(r) = q_i q_j / r$, with q_i, q_j the charges of particles $\vec{r}_i$ and $\vec{r}_j$, respectively

- A large molecules with atoms located at $\vec{r}_1, \cdots, \vec{r}_N$, with an implicit solvent. $U(\vec{r}_1, \cdots, \vec{r}_N)$ will be the usual forcefield (like that in a molecular dynamics simulations).

The Metropolis Algorithm

Given $\vec{R}^{(k)} = (\vec{r}_1^{(k)}, \cdots, \vec{r}_N^{(k)})$

- Randomly pick $J \in \{1, \cdots, N\}$ and move $\vec{r}_J^{(k)}$ to $\vec{r}_J^{(*)} = \vec{r}^{(k)} + \Delta\vec{r}$
- Define
$$\alpha = \min\left(1, \frac{\pi_\beta(\vec{R}^{(*)})}{\pi_\beta(\vec{R}^{(k)})}\right)$$
$$= \min\left(1, e^{-\beta\left[U(\vec{R}^{(*)}) - U(\vec{R}^{(k)})\right]}\right).$$

where

$$\vec{R}^* = (\vec{r}_1^{(k)}, \ldots, \vec{r}_{J-1}^{(k)}, \vec{r}_J^{(*)}, \vec{r}_{J+1}^{(k)}, \ldots, \vec{r}_N^{(k)})$$

- Accept $\vec{R}^*$ with probability α. i.e.,
  Generate $U \sim U[0,1]$.
  If $U \le \alpha$ then accept $\vec{R}^*$, and set $\vec{R}^{(k+1)} \leftarrow \vec{R}^*$.
  If $U > \alpha$ then reject $\vec{R}^*$, and set $\vec{R}^{(k+1)} \leftarrow \vec{R}^{(k)}$.

Some remarks on practical implementation of performance.

- The perturbation $\Delta\vec{r} = (\Delta x, \Delta y, \Delta z)$ is generated uniformly at random with
  $\Delta x \sim U[0, l_x]$,
  $\Delta y \sim U[0, l_y]$,
  $\Delta z \sim U[0, l_z]$,
  where $l_x, l_y, l_z > 0$ are predefined, maximal moves in x, y, z, respectively.
- A different way of moves is to move all $\vec{r}_1^{(k)}, \ldots, \vec{r}_N^{(k)}$ and then calculate the corresponding α to accept or reject these moves
- Usually the MC run consists of two stages (phases). First, the equilibriation (or

relaxation). Second, statistical analysis.

- Many systems have very complicated energy landscapes (e.g., with many local minima and saddle points). Therefore, sampling all the possible states is quite challenging in terms of efficiency. Different efficient sampling methods have been developed in recent decades. Mathematical theories for such advanced sampling techniques are still being developed.

Other topics:

- Surface growth;
- biomolecular binding and unbinding.