Chapter 3 : Monte-Carlo Method

Content

- Examples : head or tail, card games
- Historical notes
- Definition
- Sampling of a process (MCMC)
- Diffusion and random walk
- Ising Model
- Gillespie method, chemical reaction A$\rightarrow$B and $B \rightarrow A$

3.1 Motivations

- The goal of Monte-Carlo methods is the sampling of a process in order to determine some statistical properties
- For instance, we toss a coin 4 times. What is the probability to obtain 3 tail and 1 head ?
- Mathematics gives us the solution :

$$P(3 \text{ head}) = \left(\begin{array}{c} 4 \\ 3 \end{array} \right) \left(\frac{1}{2} \right)^3 \left(1 - \frac{1}{2} \right)^1 = \frac{1}{4}$$

- But we could also do a simulation

A Monte-Carlo computer simulation

```
from random import randint

success=0

attempts=10000
for i in range(attempts):
    if randint(0,1)+randint(0,1)+randint(0,1)+randint(0,1)==3:
        success+=1

print "Number of attempts=", attempts
print "Number of success=", success
```

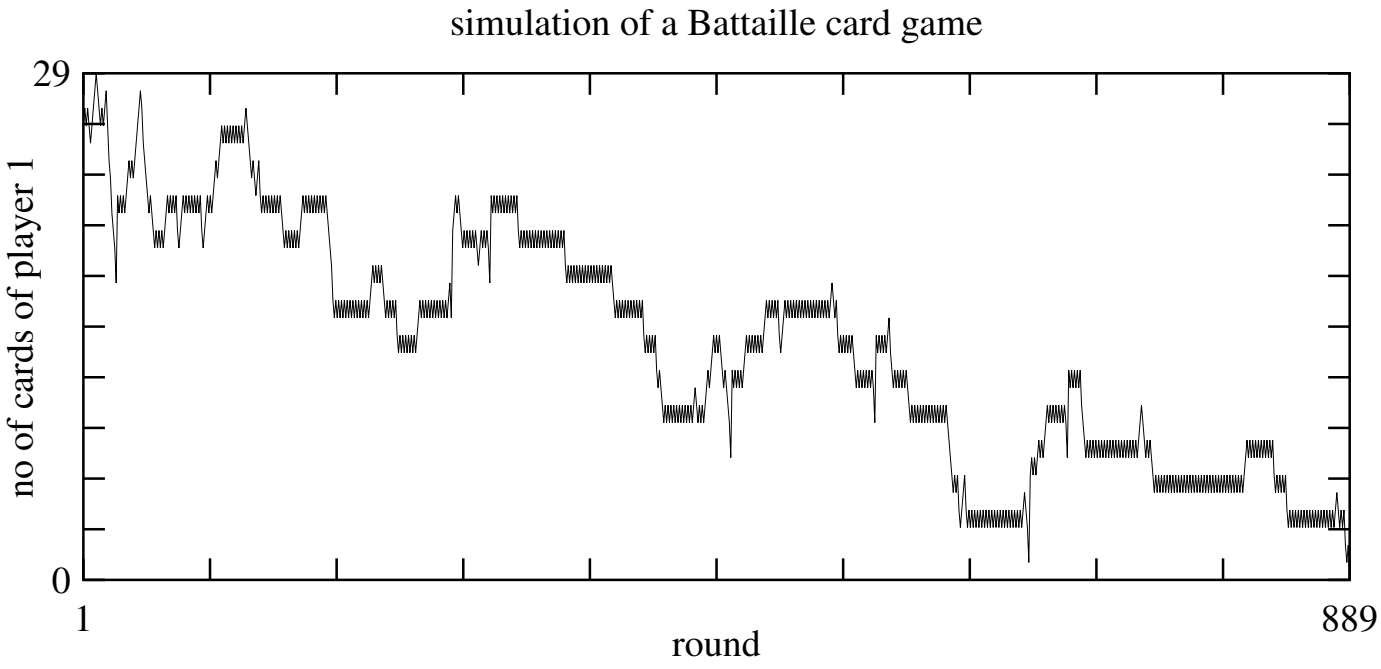
We get for instance :

```
Number of attempts= 10000
Number of success= 2559
```

More difficult problems

- For the coin tossing problem, no need for a simulation
- But we can think of other problems for which probability theory could hardly be applied
- For instance : what is the average duration of the card game called “war” (or battle) ?

The *war* card game with 52 cards



Samples of the duration of the game :
304, 86, >10000, >10000, >10000, >10000, 1110, >10000, 2315, 1904, 2570,...

Another example

- Lotka-Voltera : At each time step, one out of four process has a probability to occur : birth of a rabit/fox, death of a rabit/fox with probability that depend on the population size. (Note the *mean-field* approach)

3.2 Historical notes

- The Monte Carlo method was coined in the 1940s by **John von Neumann**, **Stanislaw Ulam** and **Nicholas Metropolis**, while they were working on nuclear weapon projects (Manhattan Project) in the Los Alamos National Laboratory.
- It was named in homage to the Monte Carlo Casino, a famous casino where Ulam's uncle would often gamble away his money.[3]

http ://en.wikipedia.org/wiki/Monte_Carlo_method

Notes historiques (Stanislaw Ulam)

The first thoughts and attempts I made to practice [the Monte Carlo Method] were suggested by a question which occurred to me in 1946 as I was playing solitaires. The question was what are the chances that a Canfield solitaire laid out with 52 cards will come out successfully ?

After spending a lot of time trying to estimate them by pure combinatorial calculations, I wondered whether a more practical method than "abstract thinking" might not be to lay it out say one hundred times and simply observe and count the number of successful plays.

Canfield Solitaire

Richard A. Canfield, owned the Canfield Casino in Saratoga Springs, New York during the 1890s. Gamblers at his casino would play the game by "buying" a deck of cards for $50. The gambler would then play the game and earn $5 for every card he managed to place into the foundations ; if one was fortunate enough to place all 52 cards into the foundations, the player would win $500.

To play the game, one must first deal thirteen cards faced up and then turned down. These cards would be the reserve, the top card of which is available for play. Then a card is placed on first of the four foundations to the right of the reserve. This card is the first card of its foundation and all other cards of the same rank must also start the other three foundations.

Source : http ://en.wikipedia.org/wiki/Canfield (solitaire)

Notes historiques (Stanislaw Ulam)

This was already possible to envisage with the beginning of the new era of fast computers, and I immediately thought of problems of neutron diffusion and other questions of mathematical physics, and more generally **how to change processes described by certain differential equations into an equivalent form interpretable as a succession of random operations**.
Later [in 1946], I described the idea to John von Neumann, and we began to plan actual calculations.

3.3 Definitions

- Monte Carlo methods (or Monte Carlo experiments) are a class of computational algorithms that rely on **repeated random sampling** to compute their results.
- Monte Carlo methods are often used in computer simulations of physical, mathematical, financial systems.
- These methods are most suited to calculations by a computer and tend to be used when it is infeasible to compute an exact result with a deterministic algorithm.
- There is a need for (pseudo) random numbers

Definitions

- There is no consensus on how Monte Carlo should be defined.
- Ripley defines most probabilistic modeling as **stochastic simulation**, with Monte Carlo being reserved for Monte Carlo integration and Monte Carlo statistical tests.
- Sawilowsky distinguishes between a **simulation**, **Monte Carlo method**, and a **Monte Carlo simulation** : a simulation is a fictitious representation of reality, a Monte Carlo method is a technique that can be used to solve a mathematical or statistical problem, and a Monte Carlo simulation uses repeated sampling to determine the properties of some phenomenon
- Monte Carlo methods are also used in the **ensemble models** that form the basis of modern weather forecasting (uncertainty quantification).

Proposed definition

A MC simulation generates, through a stochastic process, possible scenarios of evolution of a given phenomena, each with the correct probability of occurence, so as to allows one to analyze statistically the phenomena.

3.4 Sampling of a distribution (MCMC)

- The method we will discuss is often called **Markov-Chain Monte-Carlo** (MCMC)
- The goal is to sample a probability distribution $\rho(x)$ in such a large space that not all x can be taken into account
- We want to compute an average property of a system. For instance the temperature T in a gas made of many particles.
- The gas can be in so many configurations x that it is impossible to compute

$$T = \int_x T(x)\rho(x)$$

even if ρ is known.

Sampling of a distribution (MCMC)

- Instead we want to compute

$$T = \frac{1}{N} \sum_{i=1}^{N} T(x_i)$$

  where $x_1, \ldots, x_N$ are a sample of possible observations
- The challenge is to create observation points that properly sample ρ
- We can invent a dynamical process that generates the x_i with the right probability distribution.

Sampling of a distribution (MCMC)

- We consider a stochastic process in which a particle (or an observer) can jump from location x to location x' with probability $W_{x \to x'}$. This advances the system from iteration t to $t+1$ (Markov chain)
- In practice :
  1. Choose x' at random
  2. Accept x' as the new configuration if $\texttt{rand}() < W_{x \to x'}$

Sampling of a distribution (MCMC)

- We want this process to explore the space so that the particle is at point x at time t with a prescribed probability $\rho(t, x)$.
- In this way we will sample $\rho(t, x)$.
- How do we choose the $W_{x \to x'}$?

Master equation / équation maîtresse

The term "Master equation" is the name given to the equation descrbing time evolution of a probability distribution $p(t, x)$.
For a Markov process, it reads

$$p(t+1, x) = \sum_{x'} p(t, x') W_{x' \to x}$$

For us, this is the probability to find the observer at position x, at time t.

Example : random walk and diffusion equation

- One considers a 1D random walk, on a grid.
- A particle can move to the right with probability W_+, to the left with prob W_- or stay still with prob W_0.
- The master equation reads

$$p(t+1,x) = p(t,x-1)W_+ + p(t,x)W_0 + p(t,x+1)W_-$$

Example : random walk and diffusion equation

- The diffusion equation is

$$\partial_t \rho = D \partial_x^2 \rho$$

with $\rho(t, x)$ the probability to find a particle at position x at time t.

- It can be discretized as (see chapter on Dynamical Systems)

$$\rho(t+\Delta t, x) = \rho(t, x) + \frac{\Delta t D}{\Delta x^2} \left(\rho(t, x-1) - 2\rho(t, x) + \rho(t, x+1)\right)$$

Example : random walk and diffusion equation

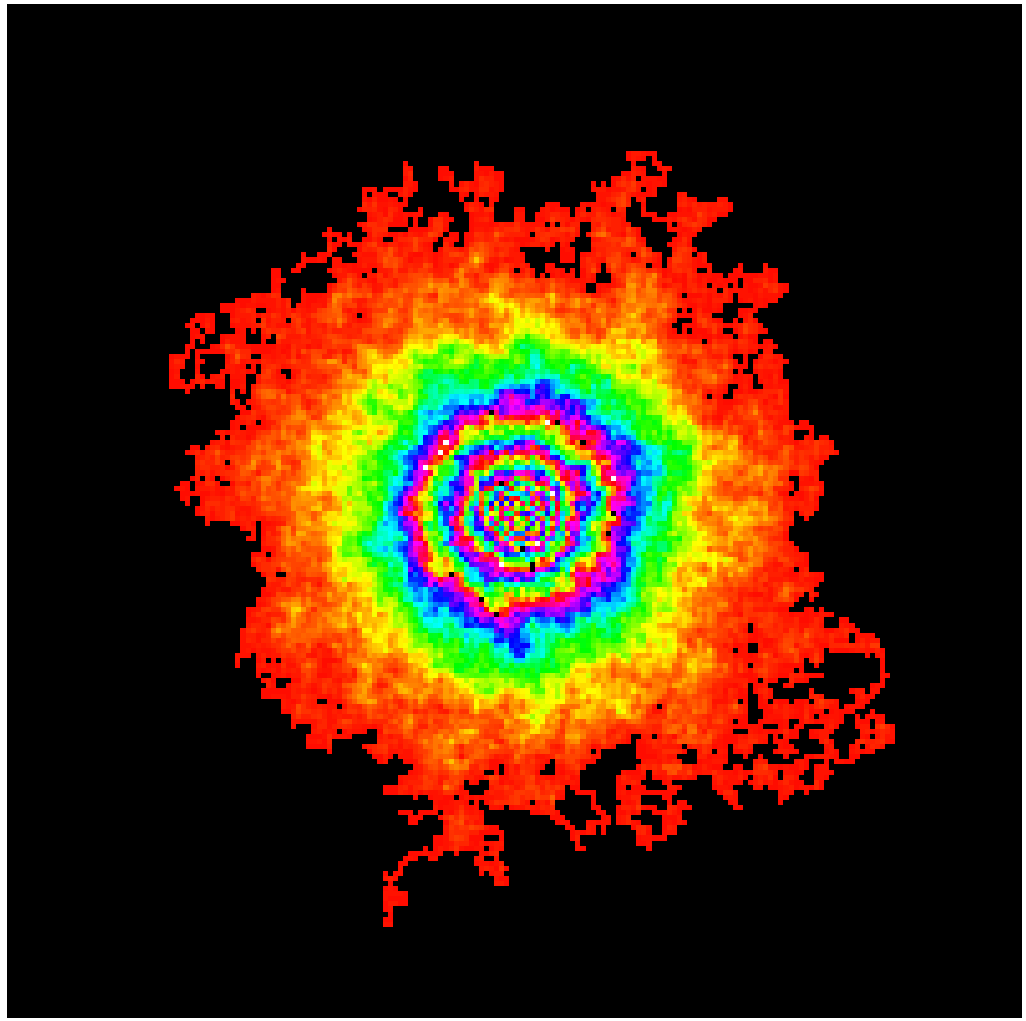
$$p(t+1,x) = p(t,x-1)W_+ + p(t,x)W_0 + p(t,x+1)W_-$$

$$\rho(t+\Delta t,x) = \rho(t,x) + \frac{\Delta t D}{\Delta x^2}\left(\rho(t,x-1) - 2\rho(t,x) + \rho(t,x+1)\right)$$

- To have $p = \rho$, one have to impose $W_+ = W_- = \Delta t D/(\Delta x)^2$ and $W_0 = 1 - 2\Delta t D/(\Delta x)^2 = 1 - W_+ - W_-$, as well as $\Delta t D/(\Delta x)^2 \leq 1/2$

In a random walk simulation it is easy to add obstacles, this may be more difficult with the differential equation (boundary conditions)

Random walk and diffusion equation



Average over 1000 particles, each moving for 1000 steps in a 200×200 lattice (`C++/RandomWalk/randomWalk.cc`).

Master equation (cont'ed)

To continue the discussion, it is useful to express the probability $p(t, x)$ to find the observing particle at location x at time t as

$$\begin{aligned} p(t+1,x) &= \sum_{x'} p(t,x') W_{x' \to x} \\ &= \sum_{x' \neq x} p(t,x') W_{x' \to x} + p(t,x) W_{x \to x} \\ &= \sum_{x' \neq x} p(t,x') W_{x' \to x} + p(t,x)(1 - \sum_{x' \neq x} W_{x \to x'}) \\ &= p(t,x) + \sum_{x' \neq x} \left[p(t,x') W_{x' \to x} - p(t,x) W_{x \to x'} \right] \end{aligned} \tag{1}$$

Detailed balance / Bilan détaillé

$$p(t+1,x) = p(t,x) + \sum_{x' \neq x} \left[p(t,x') W_{x' \to x} - p(t,x) W_{x \to x'} \right]$$

We want that $p(x)$ samples a given, distribution $\rho(x)$, independent of time, $p(t+1) = p(t)$, so that $p(t=0,x) = \rho(x)$. How to choose $W_{x \to x'}$?
The condition is

$$\sum_{x' \neq x} \left[\rho(x') W_{x' \to x} - \rho(x) W_{x \to x'} \right] = 0$$

because we want that $p(x) = \rho(x)$.

Detailed balance / Bilan détaillé

The **detailed balance** solution requires that

$$\rho(x')W_{x'\to x} - \rho(x)W_{x\to x'} = 0$$

Metropolis Rule

A way to solve the detailed balance equation is given by the so-called **Metropolis rule :**

$$W_{x \to x'} = \min\left(1, \frac{\rho(x')}{\rho(x)}\right)$$

Note that only the ratio

$$\rho(x')/\rho(x)$$

is important. No need to know the normalization of ρ.

Metropolis rule : illustration of the detailed balance

Suppose that $\rho(x') > \rho(x)$

$$\begin{aligned} \rho(x) W_{x \to x'} &= \rho(x) \min\left(1, \frac{\rho(x')}{\rho(x)}\right) \\ &= \rho(x) \\ &= \rho(x') \frac{\rho(x)}{\rho(x')} \\ &= \rho(x') W_{x' \to x} \end{aligned} \tag{2}$$

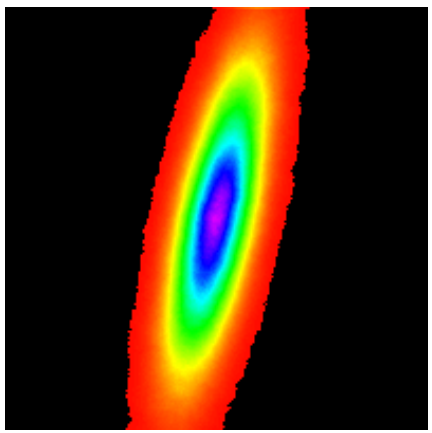
So, with the Metropolis rule, we have detailed balance

$$\rho(x') W_{x' \to x} - \rho(x) W_{x \to x'} = 0$$

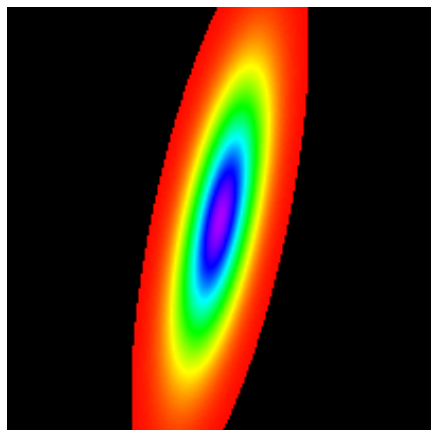
Sampling of a multivariate Gaussian distribution

Let us consider the following 2-variable Gaussian probability distribution

$$\rho(x,y) \propto \exp - \left[s_{xx}(x-\mu_x)^2 + 2s_{xy}(x-\mu_x)(y-\mu_y) + s_{yy}(y-\mu_y)^2 \right]$$



MCMC 10^8 steps



$\rho(x,y)$

The random move

$$(x,y) \to (x',y')$$

is accepted if

$$rand < \frac{\rho(x',y')}{\rho(x,y)}$$

$$\mu_x = \mu_y = 0.5 \qquad s_{xx} = 200 \qquad s_{xy} = 40 \qquad s_{yy} = 20$$

and a distretization of $[0,1]$ on a 200×200 grid.

./C++/MCMC/sampling 200 200 1000

A physical example

Let us consider a physical system at equilibrium whose probability $\rho(x)$ to be in state x is given by the Maxwell-Boltzmann distribution

$$\rho = \propto \exp(-E(x)/kT)$$

On can sample this distribution with a stochastic process by chosing $W_{x \to x'}$ with the Metropolis rule

$$W_{x \to x'} = \begin{cases} 1 & \text{si } E' < E \\ \exp[-(E' - E)/kT] & \text{si } E' > E \end{cases}$$

Consider a gas of particles : one selects one particle at random, one moves it by a given Δx, one recomputes the energy of the system with this new position (new interactions with other particles) ; then, one accepts or rejects the movement according to Metropolis rule.

Metropolis Rule

Let us assume that $E' > E$. Detailed balance is satisfied because

$$\begin{aligned}
\rho(x) W_{x \to x'} &= \Gamma \exp(-E/kT) \exp[-(E' - E)/kT] \\
&= \Gamma \exp(-E'/kT) \\
&= \rho(x') \times 1 \\
&= \rho(x') W_{x' \to x}
\end{aligned}$$

By sampling ρ with this stochastic process $W_{x \to x'}$, one can compute average physical properties (e.g. the pressure in the gas).

Glauber rule

This is another choice for $W_{x \to x'}$ given by

$$W_{x \to x'} = \frac{\rho(x')}{\rho(x) + \rho(x')}$$

which satisfies the detailed balance, too.

$$\rho(x) W_{x \to x'} = \frac{\rho(x)\rho(x')}{\rho(x) + \rho(x')} = \rho(x') W_{x' \to x}$$

Metropolis-Hastings

This is a generalization in which the choice of the candidat x' is not uniform

$$W_{x \to x'} = s_{x \to x'} w_{x \to x'}$$

where $s_{x \to x'}$ is the probability to **select** x' after x and $w_{x \to x'}$ is the probability to accept this choice.

Metropolis-Hastings

The detailed balance relation

$$\rho(x')W_{x'\to x} = \rho(x)W_{x\to x'}$$

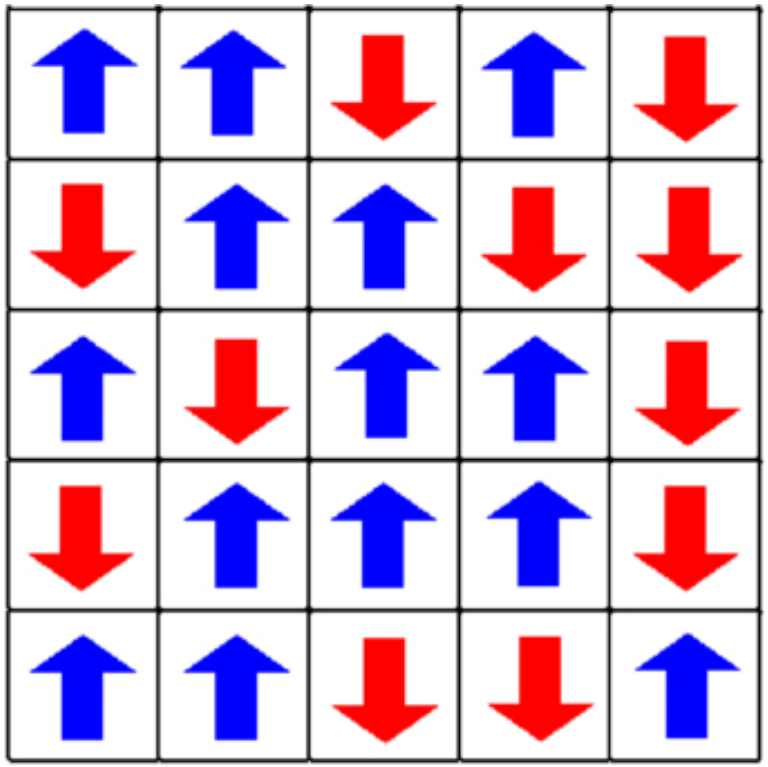
becomes

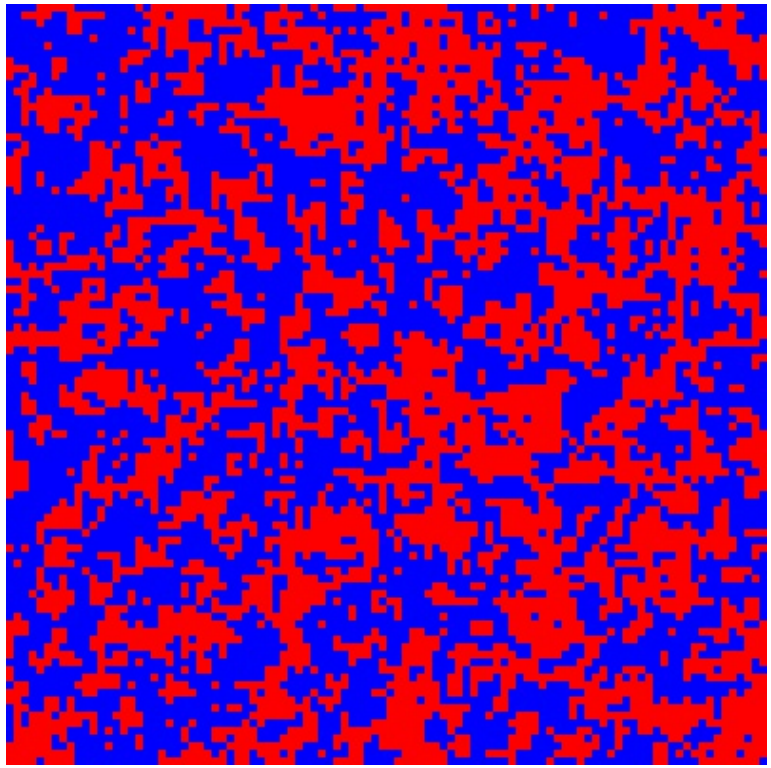
$$\frac{w_{x'\to x}}{w_{x\to x'}} = \frac{\rho(x)s_{x\to x'}}{\rho(x')s_{x'\to x}}$$

and the Metropolis solution is

$$w_{x\to x'} = \min\left(1, \frac{\rho(x')s_{x'\to x}}{\rho(x)s_{x\to x'}}\right)$$

Ising model for a magnetic material





(Images from Internet)

Ising Model

- This a model of a magnetic material. It is made of a lattice of **spins**, which are in two possible states, either *up* or `down`,
- If neighboring spins are aligned, the energy is smaller than when they are in opposit states.
- One defines $s_i \in \{-1, 1\}$ as the state of spin i
- The energy E_{ij} between two neighboring spins i et j is

$$E_{ij} = -J s_i s_j$$

- Statistical mechanics says that the probability of state $s = [s_1, s_2 \ldots, s_n]$ is given by the Maxwel-Boltzmann distribution

$$\rho(s) \sim \exp(-E/k_B T) = Z^{-1} \exp\left(-\frac{J}{k_B T} \sum_{<ij>} s_i s_j\right)$$

where $< ij >$ represents the neighboring pairs of spins.

Ising Model

- What is the behavior of this system at a given temperature T ?
- Will all the spins get aligned ?
- We will sample $\rho(s)$ to compute the average magnetization

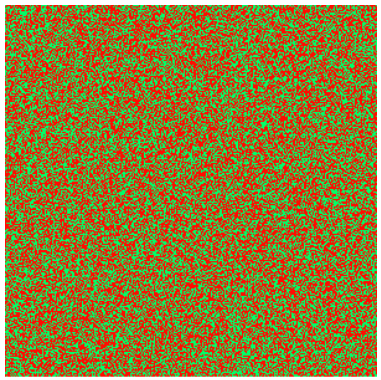
$$\langle s \rangle = \frac{1}{n} \sum_{s_1, s_2 \ldots \in \{-1,1\}} \rho(s_1, s_2, \ldots, s_n) \sum_{k=1}^{n} s_k$$

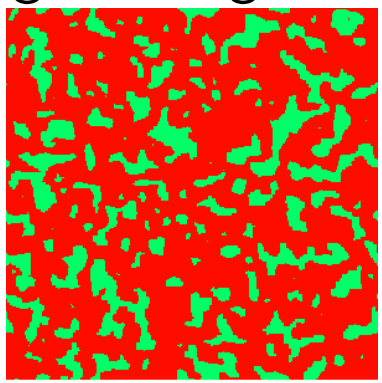
Monte-Carlo Simulation

- One chooses a spin i at random among n of them.
- The spin is flipped in state $s' = 1 - s$
- The new energy E' resulting from this flip is computed.
- The flip is accepted with probability $W_{s \to s'}$ given by Glauber or Metropolis.
- A Monte-Carlo step is done when n spins have been selected at random.
- After many such Monte-Carlo steps, the system reaches a steady state : if T is small, all the spins are aligned. If $T > T_c$ a disordered state is observed (no global magnetization).

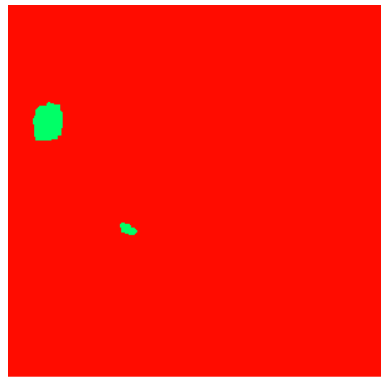
Simulations (Small temperature)

Magnetic regime $kT = 0.1$, $J = 1$


t=0


t=20


t=100


t=500

(Initial concentration of spin +1 (green) : 45%)

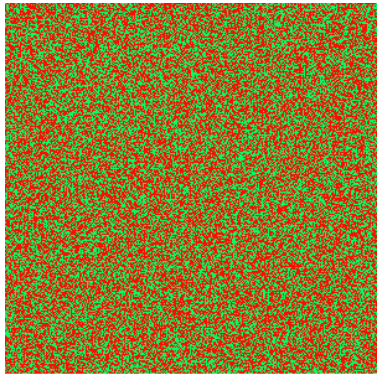
Critical temperature :

$$k_B T_c / J = \frac{2}{\ln(1 * \sqrt{2})} \approx 2.26918...$$

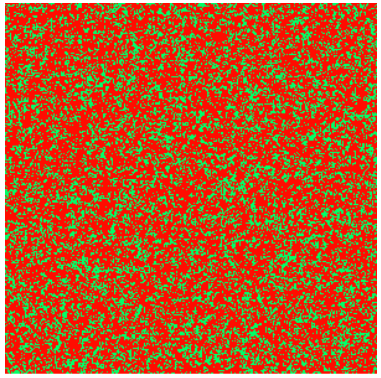
./Nextcloud/C++/Ising/ising .1 300 300 500

Simulations (high temperature)

Non-magnetic regime $kT = 5.$, $J = 1$



t=0



t=500

(Initial concentration of spin +1 (green) : 45%)

./Nextcloud/C++/Ising/ising 5. 300 300 500

3.5 Chemical reaction (Kinetic/Dynamic Monte-Carlo)

The method is illustrated on an example :
Let us consider the following reactions

$$A \xrightarrow{k_1} B \qquad B \xrightarrow{k_2} A$$

that can be written as a differential equation

$$\dot{A} = -k_1 A + k_2 B \qquad \dot{B} = +k_1 A - k_2 B$$

or, equivalently

$$\frac{d}{dt}\begin{pmatrix} A \\ B \end{pmatrix} = \begin{pmatrix} -k_1 & k_2 \\ k_1 & -k_2 \end{pmatrix}\begin{pmatrix} A \\ B \end{pmatrix}$$

Chemical reaction

This equation has an analytical solution

$$A(t) = \frac{k_2}{k_1 + k_2}(A_0 + B_0) + \frac{A_0 k_1 - B_0 k_2}{k_1 + k_2} e^{-(k_1+k_2)t}$$

$$B(t) = \frac{k_1}{k_1 + k_2}(A_0 + B_0) - \frac{A_0 k_1 - B_0 k_2}{k_1 + k_2} e^{-(k_1+k_2)t}$$

where A_0 and B_0 are the inital concentration of A et B.
If $t \to \infty$,

$$A \to A_\infty = \frac{k_2}{k_1 + k_2}(A_0 + B_0) \qquad B \to B_\infty = \frac{k_1}{k_1 + k_2}(A_0 + B_0)$$

Monte-Carlo Simulation

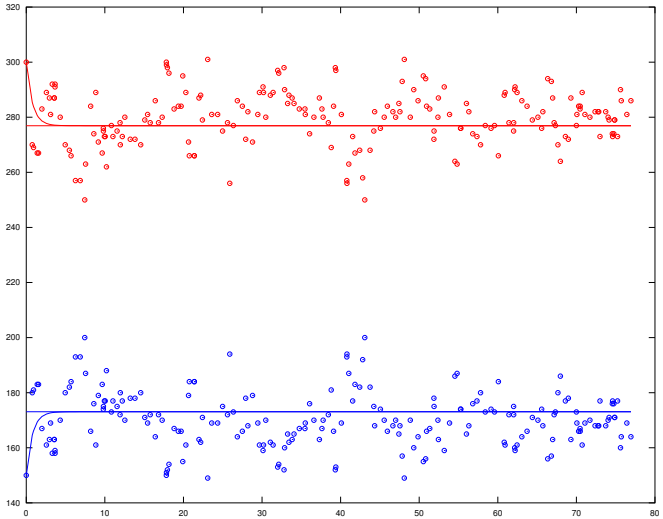
1. A time step Δt is chosen, small enough so that $k_1 \Delta t$ and $k_2 \Delta t$ are smaller than 1. The quantity $k_1 \Delta t$ is the **probability** that, during Δt, an A molecule transforms into a B molecule ; and conversely for $k_2 \Delta t$.
2. A particle is chosen at random among all $N = A(t) + B(t) = const$ particles. In practice, an A is chosen if $r < A/(A+B)$, otherwise a B is chosen, where $r = \texttt{rand}()$.

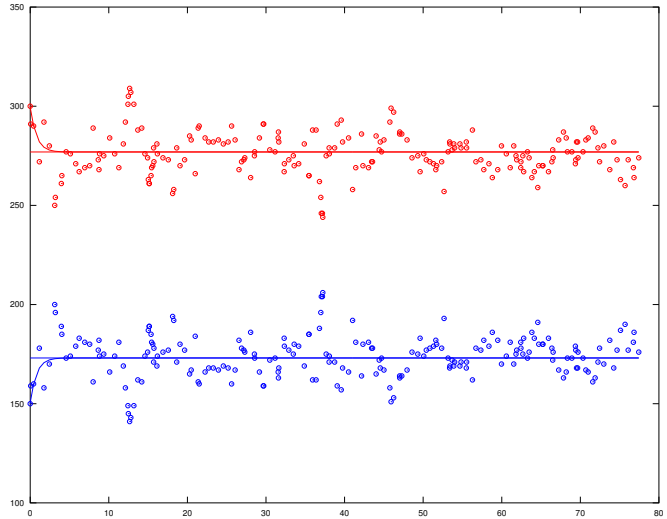
3a. If an A was chosen, it is transformed into a B provided that $r' = \texttt{rand}() < k_1 \Delta t$, Then : $A = A - 1$, $B = B + 1$.

3b. If a B was chosen, it is transformed into an A if $r' < k_2 \Delta t$. Then, $A = A + 1$, $B = B - 1$.

4. Steps (2) and (3) are repeated N times, and the time is incremented by Δt : $t = t + \Delta t$ (initially, $t = 0$).
5. Steps (2)-(4) are repeated until $t = t_{max}$

Results of two identical Monte-Carlo simulations

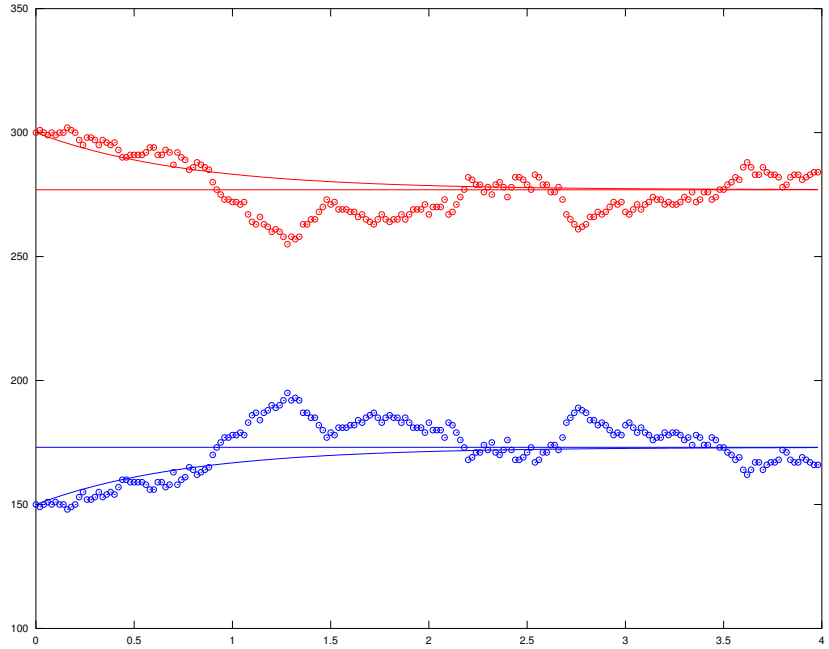




One chooses $\Delta t < 1/(k_1 + k_2)$, for instance $\Delta t = rand()/(k_1 + k_2)$. The images corresponds to $k_1 = 0.5$, $k_2 = 0.8$. Continuous lines show the analytical solutions. The points show the MC simulation. Red indicates the number of A particles, and blue shows the number of B particles. Here $A_0 = 300$, $B_0 = 150$. (Note that the scales differ in the two plots, but both runs are statistically compatibles.)

see Matlab/Gillespie/gillespie.m

Results

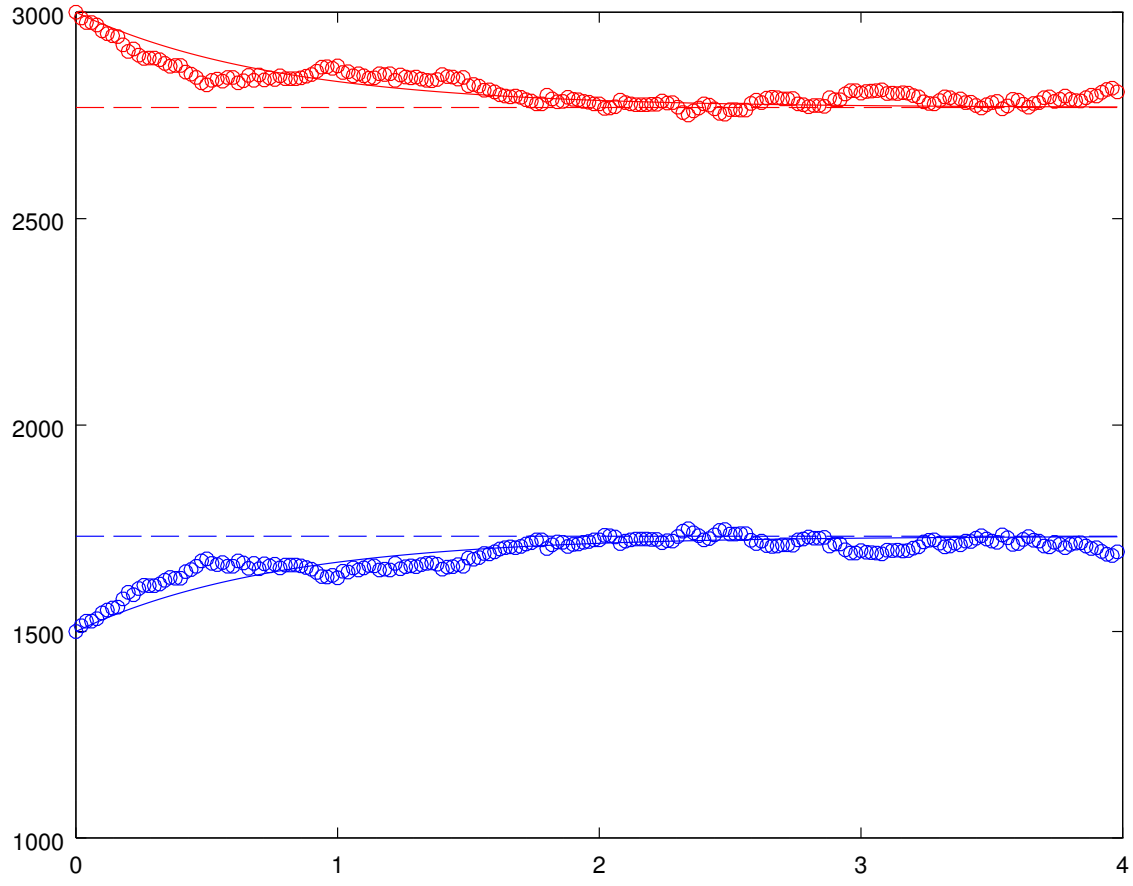


This run is with $\Delta t = 0.02$, smaller than before, and again $k_1 = 0.5$, $k_2 = 0.8$. Initial conditions : $A(t = 0) = 300$, $B(t = 0) = 150$.

MC simulations fluctuate around the analytical solution.

see Matlab/Gillespie/gillespie.m

With more particles (self average)



Same parameters but 10 times more particles : $A(t=0) = 3000$, $B(t=0) = 1500$.

Gillespie Method

This is a more elaborated algorithm

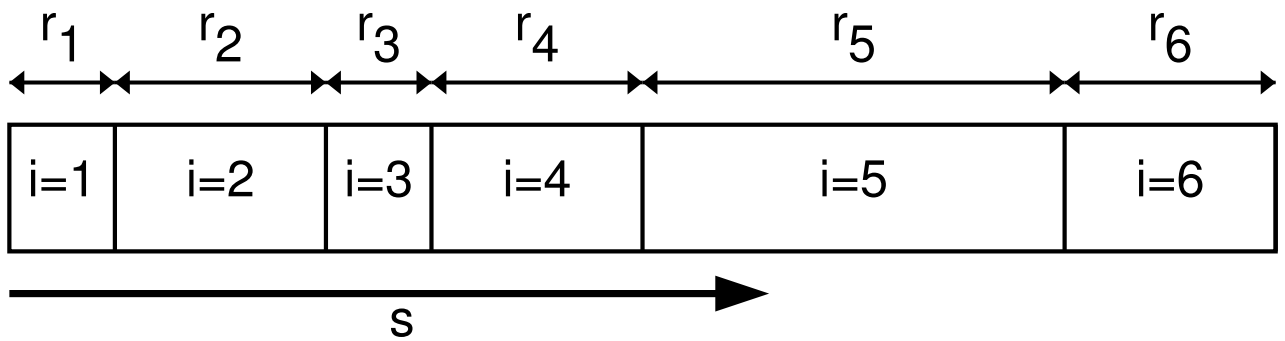
http ://en.wikipedia.org/wiki/Kinetic_Monte_Carlo

The kinetic Monte Carlo (KMC) method is a Monte Carlo method computer simulation intended to simulate the time evolution of some processes occurring in nature. Typically these are processes that occur with a given known rate. It is important to understand that these rates are inputs to the KMC algorithm, the method itself cannot predict them.

The KMC method is essentially the same as the dynamic Monte Carlo method and the Gillespie algorithm, the main difference seems to be in terminology and usage areas : KMC is used mainly in physics while the "dynamic" method is mostly used in chemistry.

Gillespie algorithm

- Let r_i be the rates at which the possible events occur in the system. $i = 1, \ldots n$.
- For instance, $r_i = kAB$ for the reaction $A + B \xrightarrow{k} C$
- Let $R_i = \sum_{j=1}^{i} r_i$ be the cumulated rates
- An event i is chosen at random by drawing a random number s : the event i which is selected is the one that verfies $R_{i-1} < sR_n < R_{i+1}$ (i.e selected with a probability proportional to rate r_i.).



Gillespie algorithm

- Realize the selected even. (For instance an A and a B react to form a C).
- Advance time by $\Delta t = \ln(1/s')/R_n$, with s' a random number.
- Note that now Δt is computed from a decreasing exponential distribution. It gives the average time of the next event : only one event can take place during Δt.