

Monte Carlo importance sampling and Markov chain

If a configuration in phase space is denoted by X , the probability for configuration according to Boltzman is

$$\rho(X) \propto e^{-\beta E(X)} \quad \beta = \frac{1}{T} \quad (1)$$

How to sample over the whole phase space for a general problem? How to generate configurations?

- **Brute force:** generate a truly random configuration X and accept it with probability $e^{-\beta E(X)}$ where all $E > 0$. Successive X are **statistically independent**. *VERY INEFFICIENT*
- **Markov chain:** Successive configurations X_i, X_{i+1} are **NOT** statistically independent but are distributed according to Boltzman distribution.

What is the difference between Markov chain and uncorrelated sequence?

- *Truly random or uncorrelated* sequence of configurations satisfies the identity

$$P(X_1, X_2, \dots, X_N) = P_1(X_1)P_1(X_2) \cdots P_1(X_N)$$

- *Markov chain* satisfies the equation

$$P(X_1, X_2, \dots, P_{X_N}) = P_1(X_1)T(X_1 \rightarrow X_2)T(X_2 \rightarrow X_3) \cdots T(X_{N-1} \rightarrow X_N)$$

where the transition probabilities $T(X \rightarrow X')$ are normalized

$$\sum_{X'} T(X \rightarrow X') = 1$$

We want to generate Markov chain where distribution of states is proportional to $e^{-\beta E(X)}$ and this distribution should be independent of the position within the chain and independent of the initial configuration.

The necessary conditions for generating such Markov chain is that every configuration in phase space should be accesible from any other configuration within finite number of steps (*connectedness* or *irreducibility*) - (Be careful to check this condition when choosing Monte Carlo step!)

We need to find transition probability $T(X \rightarrow X')$ which leads to a given **stationary** distribution $\rho(X)$ (in this case $\rho(X) \propto e^{-\beta E(X)}$).

The probability for X decreases, if system goes from X to any other X' :

– $\sum_{X'} \rho(X)T(X \rightarrow X')$ and increases if X configuration is visited from any other state X' : $\sum_{X'} \rho(X')T(X' \rightarrow X)$. The change of probability for X is therefore

$$\rho(X, t + 1) - \rho(X, t) = - \sum_{X'} \rho(X)T(X \rightarrow X') + \sum_{X'} \rho(X')T(X' \rightarrow X) \quad (2)$$

We look for stationary solution, i.e., $\rho(X, t + 1) - \rho(X, t) = 0$ and therefore

$$\sum_{X'} \rho(X)T(X \rightarrow X') = \sum_{X'} \rho(X')T(X' \rightarrow X) \quad (3)$$

General solution of this equation is not accesible, but a particular solution is obvious

$$\rho(X)T(X \rightarrow X') = \rho(X')T(X' \rightarrow X) \quad (4)$$

This solution is called **DETAIL BALANCE** solution.

To construct algorithm, we define transition prob. $T(X \rightarrow X') = \omega_{XX'} A_{XX'}$:

- *trial step probability* $\omega_{XX'}$ which is symmetric, i.e., $\omega_{XX'} = \omega_{X'X}$ (for example spin flip in ising: $\omega_{XX'}$ is $1/L^2$ if X and X' differ for a single spin flip and zero otherwise) and
- *acceptance probability* $A_{XX'}$ (for example accepting or rejecting new configuration with probability proportional to $\min(1, \exp(-\beta(E(X') - E(X))))$).

Detail balance condition becomes

$$\frac{\rho(X')}{\rho(X)} = \frac{A_{XX'}}{A_{X'X}}$$

Metropolis chooses

$$\begin{aligned} A_{XX'} &= 1 && \text{if } \rho(X') > \rho(X) \\ A_{XX'} &= \frac{\rho(X')}{\rho(X)} && \text{if } \rho(X') < \rho(X). \end{aligned} \tag{5}$$

Obviously, this acceptance probability satisfies detail balance condition and therefore leads to desired Markov chain with stationary probability for any configuration $X \propto \rho(X)$ for long times.

To summarize Metropolis algorithm

- $T(X \rightarrow X') = \omega_{XX'} A_{XX'}$
- $\sum_{X'} \omega_{XX'} = 1; \omega_{XX'} = \omega_{X'X}$
- $\omega_{XX'} > 0$ for all X, X' after finite number of steps
- $A_{XX'} = \min(1, \frac{\rho(X')}{\rho(X)})$

How to accept a step with probability $A_{XX'} < 1$? One can generate a random number $r \in [0, 1]$ and accept the step if $r < A_{XX'}$.

Keep in mind:

- Configurations that are generated by Markov chain are correlated. The theory guarantees that we arrive at invariant distribution ρ for long times.
- Two configurations are statistically independent only if they are far apart in the Markov chain. This distance is called *correlation time* (*Be careful: To measure distance in Markov chain, every step counts, not only successful.*)

The average of any quantity can be calculated as usual

$$\bar{A} = \frac{1}{n - n_0} \sum_{i > n_0}^n A_i$$

where n_0 steps are used to "warm-up".

The error of the quantity, however, is much bigger than the following quantity

$$\frac{1}{n - n_0} \sum_{i > n_0}^n (A_i - \bar{A})^2$$

Imagine the extreme limit of correlations when all values A_i are the same. We would estimate that standard deviation is zero regardless of the actual error!

To compute standard deviation, we need to group measurements within the correlation time into bins and then estimate the standard deviation of the bins:

$$B_l = \frac{1}{N_0} \sum_{i=N_l}^{i < N_l + N_0} A_i \quad (6)$$

$$\sigma^2 = \frac{1}{M} \sum_{j=0}^{M-1} (B_j - \bar{A})^2 \quad (7)$$

where we took into account that $\bar{A} = \bar{B}$. The correlation time (here denoted by N_0) is not very easy to estimate. Maybe the best algorithm is to compute σ^2 for few different N_0 and as long as σ^2 is increasing with N_0 , the correlation time is still larger than N_0 . When σ^2 stops changing with increasing N_0 , we reached correlation time and σ^2 is a good estimation of standard deviation.
