Indian Journal of Economics and Development, Vol 6(2)February 2018 **ISSN(online):2320-9836**
ISSN(Print):2320-9828

Metropolis and Wang-Landau Algorithms

T. Pramananda Perumal*, K. P. N. Murthy†

* Presidency College, Kamarajar Salai, Triplicane, Chennai 600 005, Tamilnadu, India
† Chennai Mathematical Institute (CMI), H1, SIPCOT IT Park, Siruseri, Kelambakkam, Chennai 603 103, Tamilnadu, India
pramanandaperumal@yahoo.com, k.p.n.murthy@gmail.com

Abstract

Objectives: We review two algorithms developed for simulating macroscopic systems. The first is the Metropolis and the second is the Wang-Landau algorithm.
Methods: Metropolis algorithm has been extensively employed for simulating a canonical ensemble and estimating macroscopic properties of a closed system at any desired temperature. A mechanical property, like energy can be calculated by averaging over a large number of micro states of the stationary Markov chain generated by the Metropolis algorithm. However thermal properties like entropy, and free energies are not easily accessible. A method called *umbrella sampling* was proposed some forty years ago for this purpose. Ever since, umbrella sampling has undergone several metamorphoses and we have now multi canonical Monte Carlo, entropic sampling, flat histogram methods, Wang-Landau algorithm *etc.*
Findings: In this paper we review Metropolis algorithm for estimating mechanical properties and Wang-Landau algorithm for estimating both mechanical and thermal properties of an equilibrium system.
Applications: We shall make the review as pedagogical and as self-contained as possible. These algorithms can be applied to a variety of problems in physics, astrophysics, chemistry, biology, soft matter, computer science, *etc.*
Keywords **:** Monte Carlo simulation; Metropolis algorithm; entropic sampling; flat-histogram methods; detailed balance; Markov Chain.

1 Some Preliminaries

The subject of statistical mechanics helps us go from the micro world of atoms and molecules obeying laws of classical and quantum mechanics to the macro world of thermodynamics, describing matter in bulk. In one single stroke Ludwig Eduard Boltzmann (1844 - 1906) connected physics at the length scales of atoms and molecules to phenomenon happening on on the length scales of solids, liquids, gases, polymers, magnets, *etc.* The micro-macro synthesis proceeds, very generally along the following lines.

First, we identify a random variable that corresponds to a thermodynamic property. We shall be concerned with equilibrium systems only. The average of the random variable over a suitable and well defined statistical ensemble[1] gives the value of the thermodynamic property. As an example, consider internal energy[2] of a thermodynamic system. This property is usually denoted by the symbol $\boldsymbol{U}$.

[1] *e.g.* micro canonical ensemble for an isolated system; canonical ensemble for a closed system; and grand canonical ensemble for an open system.

[2] see Appendix - A on Internal Energy and the First Law of Thermodynamics

Corresponding to this property, we have, in statistical mechanics, energy $\boldsymbol{E}$ - the kinetic energy and the interaction energy of the atoms and molecules of the macroscopic object. A numerical value for $\boldsymbol{E}$ can be assigned to each micro state[3] of the macroscopic system. The value of $\boldsymbol{E}$ will fluctuate when the equilibrium system goes from one micro state to another. These fluctuations are an integral part of an equilibrium descriptions. The average of $\boldsymbol{E}$ gives the internal energy : $\langle \boldsymbol{E} \rangle = \boldsymbol{U}$, and the fluctuations give the heat capacity :

$$\langle \boldsymbol{E^2} \rangle - \langle \boldsymbol{E} \rangle^{\boldsymbol{2}} = \boldsymbol{k_B T^2 C_V}.$$

The symbol $\langle (\cdot) \rangle$ denotes averaging of the property $(\cdot)$ over a canonical ensemble.

Energy

The computation of average energy is now straight forward : Generate a canonical ensemble employing, for example, a Monte Carlo method based on Metropolis algorithm; a simple arithmetic average of $\boldsymbol{E}$ over a Monte Carlo sample of reasonably large size, gives the required answer. The statistical error associated with the finite-sample average can also be calculated from the data obtained in the simulation. Such a neat computational scheme is possible because a numerical value for energy can be assigned to each micro state of the macroscopic system.

Entropy

How does one calculate entropy ?

We can not assign a numerical value for entropy to any single micro state. Entropy is a property that belongs collectively to all the micro states. While energy is a *private* property (of each micro state), entropy is a *social* or a *pubic* property, see below.

Let $\boldsymbol{\Omega}$ denote the set of micro states of an equilibrium system; the micro states are discrete, distinct and finite in number. Let $\widehat{\boldsymbol{\Omega}}$ denote the number of micro states in the set $\boldsymbol{\Omega}$. Formally, let $\{\boldsymbol{p}(\boldsymbol{\mathcal{C}_\nu}) \ : \ \boldsymbol{\nu = 1,\ 2,\ \cdots, \widehat{\Omega}}\}$ denote the probabilities of the micro states. We use script $\boldsymbol{\mathcal{C}}$ to denote micro states of the system and regular Roman $\mathbf{C}$ to denote micro states of an ensemble or of a Monte Carlo sample or of a Markov chain. The Boltzmann-Gibbs-Shannon entropy is given by

$$\boldsymbol{S = -k_B \sum_{\nu=1}^{\widehat{\Omega}} p(\mathcal{C}_\nu)\ \ln p(\mathcal{C}_\nu).}$$

In the above, $\boldsymbol{k_B}$ is the Boltzmann constant[4].

Entropy of an Isolated System

For an isolated system, the micro states are equally probable[5]. We have,

$$\boldsymbol{p(\mathcal{C}_\nu) = \frac{1}{\widehat{\Omega}(E, V, N)} \ \forall\ \nu.}$$

[3]For example three positions $(\boldsymbol{q_1, q_2, q_3})$ and three momenta $(\boldsymbol{p_1, p_2, p_3})$ are required to specify a single point particle. For $\boldsymbol{N}$ particles, we need a string of $\boldsymbol{6N}$ numbers and this string denotes a micro state of the macroscopic system of $\boldsymbol{N}$ particles. $\boldsymbol{E = (1/2m) \sum_{i=1}^{3N} p_i^2 + V(q_1, q_2, \cdots, q_{3N})}$ - the kinetic and potential energies. Note that energy is defined for each micro state. For a macroscopic system of say $\boldsymbol{N}$ Ising spins, we have $\boldsymbol{2^N}$ micro states; note each Ising spin can be in either "up" $(\boldsymbol{S_i = +1})$ state or "down" $(\boldsymbol{S_i = -1})$ state. $\boldsymbol{E = -J \sum_{\langle i,j \rangle} S_i\ S_j}$ where $\boldsymbol{S_i}$ is the spin at lattice site $\boldsymbol{i}$ and $\boldsymbol{J > 0}$ measures the strength of spin-spin interaction. Spins on nearest neighbour lattice sites interact. The sum runs over all pairs of nearest neighbour spins

[4]$\boldsymbol{k_B = 1.38064852 \times 10^{-23}}$ joules/kelvin

[5]we call it *ergodicity*; it is an hypothesis; the entire edifice of statistical mechanics is built on this hypothesis.

$\widehat{\Omega}(E, V, N)$ is the number of micro states of the isolated system of of N particles, confined to a volume V, and with a fixed total energy of E. For an isolated system the expression for entropy simplifies to

$$S(E, V, N) = k_B \ln \widehat{\Omega}(E, V, N).$$

Entropy of a Closed System

For a closed system at temperature[6] $T = 1/[k_B\beta]$, we have

$$p(\mathcal{C}_\nu) = \frac{1}{Q} \exp[-\beta E(\mathcal{C}_\nu)],$$

where

$$Q = \sum_{\nu=1}^{\widehat{\Omega}} \exp[-\beta E(\mathcal{C}_\nu)]$$

is called the canonical partition function.

Entropy of an Open System

For an open system we have

$$p(\mathcal{C}_\nu) = \frac{1}{\mathcal{Q}} \exp[-\beta\{E(\mathcal{C}_\nu) - \mu N(\mathcal{C}_\nu)\}],$$

where μ is the chemical potential[7] of the system, $N(\mathcal{C}_\nu)$ is the number of particles in the system when it is in micro state $\mathcal{C}_\nu$, and

$$\mathcal{Q}(T, V, \mu) = \sum_{\nu=1}^{\widehat{\Omega}} \exp[-\beta\{E(\mathcal{C}_\nu) - \mu N(\mathcal{C}_\nu)\}],$$

is the grand canonical partition function.

Our aim is to simulate physical processes occurring in an equilibrium system and assemble a large ensemble of micro states consistent with the given probabilities. To this end, we start with an arbitrary initial micro state $\mathbf{C_0}(\in \Omega)$ and generate a Markov chain[8]

$$\mathbf{C_0}(\in \Omega) \to \mathbf{C_1}(\in \Omega) \to \mathbf{C_2}(\in \Omega) \to \cdots \to \mathbf{C}_i(\in \Omega) \to \mathbf{C}_{i+1}(\in \Omega) \to \cdots$$

employing Metropolis rejection algorithm [1], see the next section.

Note we have used Roman $\mathbf{C}$ to denote the micro states in the Markov chain. The subscript stands for discrete time. Each micro state in the Markov chain is drawn from the set Ω containing the micro states of the system. In the chain, it is quite possible $\mathbf{C}_i = \mathbf{C}_j$ for several i s and j s.

---

[6]In thermodynamics, temperature is defined as $T = \left(\frac{\partial U}{\partial S}\right)_{V,N}$.

[7]the chemical potential gives the change in energy upon addition of a single particle keeping the entropy and volume of the system at a constant value. In other words $\mu = \left(\frac{\partial U}{\partial N}\right)_{S,V}$.

[8]See Appendix-B on Markov Chain

2 Metropolis Rejection Algorithm

Let us say we have simulated the Markov chain upto $\mathbf{C}_i \in \mathbf{\Omega}$ starting from $\mathbf{C_0} \in \mathbf{\Omega}$. Thus the current micro state is $\mathbf{C}_i$. Let $\boldsymbol{p_i} = \boldsymbol{p}(\mathbf{C}_i)$ denote its probability. We make a small random change in the current micro state and construct a trial micro state[9] $\mathbf{C}_t \in \mathbf{\Omega}$. Let $\boldsymbol{p_t} = \boldsymbol{p}(\mathbf{C}_t)$ denote its probability. Calculate $\boldsymbol{p} = \textbf{minimum}\,(\mathbf{1},\ \boldsymbol{p_t/p_i})$ **.** Then, the next micro state in the Markov chain is given by,

$$\mathbf{C}_{i+1} = \begin{cases} \mathbf{C}_t & \textbf{with probability} \quad \boldsymbol{p} \\ \mathbf{C}_i & \textbf{with probability} \quad \mathbf{1} - \boldsymbol{p} \end{cases}$$

The implementation of the above goes as follows :

- Generate a random number[10] uniformly distributed between zero and unity. Denote it by the symbol $\boldsymbol{\xi}$.
- If $\boldsymbol{\xi} \ \leq \ \boldsymbol{p}$, then accept the trial state and advance the Markov chain to $\mathbf{C}_{i+1} = \mathbf{C}_t$.
- If not, reject the trial state and advance the Markov chain to $\mathbf{C}_{i+1} = \mathbf{C}_i$.
- Repeat the process on $\mathbf{C}_{i+1}$ to get $\mathbf{C}_{i+2}$; and so on.

Generate a long Markov chain. The asymptotic part of the chain shall contain micro states belonging to the ensemble characterized by the probabilities $\{\boldsymbol{p}(\mathbf{C}_\nu)\ :\ \boldsymbol{\nu} = \mathbf{1},\ \mathbf{2},\ \cdots\}$.

Important Features of the Metropolis Algorithm

- The Metropolis algorithm demands only a knowledge of the ratio of probabilities of any two micro states belonging to $\mathbf{\Omega}$. This is the most important property of the Metropolis algorithm. As a conseqquence of this property, we need to know $\{\boldsymbol{p}(\boldsymbol{\mathcal{C}}_\nu) : \boldsymbol{\nu} = \mathbf{1}, \mathbf{2}, \cdots \widehat{\mathbf{\Omega}}\}$ only up to a normalization constant. It is precisely because of this reason we are able to simulate a closed system, since we need to know only the Boltzmann weight $\mathbf{exp}[-\boldsymbol{\beta E}(\boldsymbol{\mathcal{C}})]$ of each micro state; we need not have any knowledge what so ever of the canonical partition function.
- The next important property is that the Metropolis algorithm obeys balance condition[11]. This ensures that the Markov chain shall definitely converge to an invariant probability density (of the micro states).
- Another important property is that the Metropolis algorithm obeys a stricter condition called detailed balance[12]

  The consequences of this are two fold.

[9]For example if we are simulating an Ising spin system, select randomly an Ising spin from the current spin configuration (micro state) and flip it to construct a trial spin configuration. If we are simulating a collection of particles, then select a particle randomly and change its there position coordinates and three momentum coordinates by small random amounts to construct a trial micro state.

[10]employ the random number generator available in your computer. The (pseudo) random numbers are real numbers independently and uniformly distributed between zero and one. Random number generation and testing are non-trivial tasks and they constitute highly specialized areas of research. Most Monte Carlo practitioners are not aware of the subtleties and difficulties associated with random number generation employing deterministic algorithms and testing of the generated random numbers for randomness. We take the available random generator and use it as a black box.

[11]see Appendix-C on Time Homogeneous Markov Chain.

[12]see section under Detailed Balance in Appendix - C on Time Homogeneous Markov Chain

(i) Detailed balance ensures the Markov chain converges to an equilibrium ensemble consistent with the given probability weights of the micro states : Boltzmann weights for canonical ensemble; and Gibbs weights for grand canonical ensemble; *etc.* We can choose the nature of the equilibrium state.

(ii) Detailed balance ensures the Markov chain is reversible, see *e.g.* [2], and hence is most suited for describing an equilibrium system[13]; for, no matter what kind of observations you make on an equilibrium system, you can not tell which way time moves, past to present to future or the reverse, future to present to past. Equilibrium is a time-reversal invariant state. Detailed balance captures this subtle property. Note also that detailed balance implies balance and not *vice versa*.

Estimation of Averages and Statistical Errors

Generate a Markov chain until it equilibrates[14]. Continue the Markov chain and collect a reasonably large number of micro states $\{\mathbf{C}_i \ : \ i = 1, \ 2, \ \cdots M\}$ from the equilibrated Markov chain. Let $\boldsymbol{O}$ be a property of interest and $\boldsymbol{O(C)}$ its value when the system is in micro state $\boldsymbol{C}$. Then the Monte Carlo estimate of the property $\boldsymbol{O}$ is given by[15],

$$\overline{\boldsymbol{O}}_{\boldsymbol{M}} = \frac{\mathbf{1}}{\boldsymbol{M}} \sum_{\boldsymbol{i=1}}^{\boldsymbol{M}} \boldsymbol{O}(\mathbf{C}_i); \qquad \underset{\boldsymbol{M\to\infty}}{\textbf{Limit}} \ \overline{\boldsymbol{O}}_{\boldsymbol{M}} = \langle \boldsymbol{O} \rangle .$$

A little thought will tell you that the quantity $\overline{\boldsymbol{O}}_{\boldsymbol{M}}$ is a random variable. It will fluctuate from one realization of a Monte Carlo sample to another.

What is the nature of these fluctuations ?

The Central limit theorem[16] (CLT) tells that the quantity $\overline{\boldsymbol{O}}_{\boldsymbol{M}}$ is a Gaussian random variable when $\boldsymbol{M}$ is large. The average of the Gaussian is $\langle \boldsymbol{O} \rangle$ and its variance is $\boldsymbol{\sigma^2/M}$, where $\boldsymbol{\sigma^2} = \langle \boldsymbol{O^2} \rangle - \langle \boldsymbol{O} \rangle^{\boldsymbol{2}}$. A possible statement of the statistical error associated with the Monte Carlo estimate $\overline{\boldsymbol{O}}_{\boldsymbol{M}}$ is obtained from the following considerations.

Take a Gaussian random variable with mean $\boldsymbol{\zeta}$ and standard deviation $\boldsymbol{\Sigma}$. The area under the Gaussian[17] between $\boldsymbol{\zeta - \Sigma}$ and $\boldsymbol{\zeta + \Sigma}$ is $\mathbf{0.682695}$. This means that with $\mathbf{68.27\%}$ confidence, you can say that a randomly sampled number from the Gaussian shall lie between $\boldsymbol{\zeta - \sigma}$ and $\boldsymbol{\zeta + \sigma}$. The one-sigma confidence interval provides a neat quantification of the statistical error associated with Monte Carlo estimates, see below.

We calculate the second moment,

$$\overline{\boldsymbol{O}}^{\boldsymbol{2}}_{\boldsymbol{M}} = \frac{\mathbf{1}}{\boldsymbol{M}} \sum_{\boldsymbol{i=1}}^{\boldsymbol{M}} \boldsymbol{O^2}(\mathbf{C}_i); \qquad \underset{\boldsymbol{M\to\infty}}{\textbf{Limit}} \ \overline{\boldsymbol{O}}^{\boldsymbol{2}}_{\boldsymbol{M}} = \langle \boldsymbol{O^2} \rangle .$$

[13] see Appendix-D on Time Symmetry

[14] calculate the moving average of energy. When it stabilizes to a constant value but for some small statistical fluctuations, we can say the system has equilibrated.

[15] We reserve the symbol $\langle \boldsymbol{O} \rangle$ to denote the unknown exact value of the canonical ensemble average of the property $\boldsymbol{O}$ formally given by

$$\langle \boldsymbol{O} \rangle = \frac{\mathbf{1}}{\boldsymbol{Q}} \sum_{\boldsymbol{\nu=1}}^{\hat{\boldsymbol{\Omega}}} \boldsymbol{O}(\mathcal{C}_\nu) \exp[-\boldsymbol{\beta E}(\mathcal{C}_\nu)]; \qquad \boldsymbol{Q} = \sum_{\boldsymbol{\nu=1}}^{\hat{\boldsymbol{\Omega}}} \exp[-\boldsymbol{\beta E}(\mathcal{C}_\nu)].$$

[16] see Appendix-E on Central Limit Theorem

[17]
$$\frac{\mathbf{1}}{\boldsymbol{\Sigma}\sqrt{\mathbf{2}\boldsymbol{\pi}}} \int_{\boldsymbol{\zeta-\Sigma}}^{\boldsymbol{\zeta+\Sigma}} \boldsymbol{dx} \ \exp\left[-\frac{(\boldsymbol{x}-\boldsymbol{\zeta})^{\mathbf{2}}}{\mathbf{2}\boldsymbol{\Sigma}^{\mathbf{2}}}\right] = \mathbf{0.682695}$$

From the calculated values of the first and second moments we estimate the variance as,

$$\boldsymbol{\sigma^2_M = \overline{O}^2_M - (\overline{O}_M)^2.} \quad \left(\boldsymbol{\sigma^2 = \langle O^2\rangle - \langle O\rangle^2 = \underset{M\to\infty}{\text{Limit}}\ \sigma^2_M}\right).$$

We then estimate the one-sigma confidence interval and quote the Monte Carlo result as $\boldsymbol{\overline{O}_M \pm \sigma_M/\sqrt{M}}$**.** The above means that with **0.6827** probability we can expect the Monte Carlo estimate $\boldsymbol{\overline{O}_M}$ to lie in the one sigma interval around $\boldsymbol{\langle O\rangle}$, *i.e.* to lie between $\boldsymbol{\langle O\rangle - \sigma_M/\sqrt{M}}$ and $\boldsymbol{\langle O\rangle + \sigma_M/\sqrt{M}}$.

The statistical error decreases with increase of $\boldsymbol{M}$. This is indeed a desirable property. This tells us, atleast in principle, we will get things right if $\boldsymbol{M}$ is sufficiently large. Usually we would be interested in comparing our Monte Carlo predictions with experiments. Hence we can take the Monte Carlo sample size to be large enough to ensure that the statistical error is less that the experimental error bar.

However, notice the statistical error decreases painfully slowly with the sample size. The decrease is logarithmically slow : to better the results by one extra decimal accuracy we need to increase the sample size a hundred fold. Often this would prove to be an exercise in futility; for, the computing time is linear in $\boldsymbol{M}$.

We need variance reduction devices that significantly reduce the fluctuations without affecting the averages. Importance sampling is a variance reduction device. It helps us sample micro states from important regions of the sample space *e.g.* micro states with high Boltzmann weights. Notice a randomly selected micro state would be, most likely, of high energy[18], hence of low Boltzmann weight; its contribution to the partition sum would be negligible. In fact the Metropolis algorithm is an importance sampling device. I am not going to talk of importance sampling or of other variance reduction techniques; those interested can consult for example [5, 6, 7].

Instead, in what follows, I am going to investigate the nature of the invariant distribution of the Markov chain of micro states whose probabilities are inversely proportional to the density of states : micro states of high entropy region have low probabilities; and those of low entropy region have high probabilities. This kind of prescription does not help simulate any physical system or any physical processes. Nevertheless it has certain advantages and this will become clear in the sequel.

3 Markov Chain with Flat Energy Histogram

Consider a system with micro states $\boldsymbol{\Omega = \{\mathcal{C}_\nu\ :\ \nu = 1, 2, \cdots, M\}}$. Let $\boldsymbol{\widehat{\Omega}(E)}$ denote its density of states. For purpose of illustration we assume that the density of states is known. Let $\boldsymbol{\mathcal{C}_\mu \in \Omega}$ and $\boldsymbol{E_\mu = E(\mathcal{C}_\mu)}$. We prescribe

$$\boldsymbol{P(\mathcal{C}_\mu)\ \propto\ \frac{1}{\widehat{\Omega}(E\mu)}.}$$

Let me emphasize two points before we proceed further :

- We do not know the density of states. After all, if we know the density of states we can estimate all the properties of the system employing the well developed machinery of thermodynamics and statistical mechanics. We do not need Monte Carlo simulation.

- There is no physical system for which the probability of a micro state is inversely proportional to the density of states.

[18] entropy increases with energy.

Nevertheless we shall consider Monte Carlo simulation of such an un-physical system employing Metropolis algorithm and investigate the nature of the micro states of the equilibrated Markov chain. Accordingly we generate a Markov chain based on these probabilities and Metropolis rule. Note we need to know the probabilities of the micro states only upto a normalization constant.

Let $\mathbf{C}_i$ be the current micro state in the Markov chain and $\boldsymbol{E_i} = \boldsymbol{E}(\mathbf{C}_i)$ its energy. We take

$$\boldsymbol{p_i} = \boldsymbol{p}(\mathbf{C}_i) \propto \frac{\mathbf{1}}{\widehat{\boldsymbol{\Omega}}(\boldsymbol{E_i})},$$

as its probability. Let $\mathbf{C}_t$ be the trial state and $\boldsymbol{E_t} = \boldsymbol{E}(\mathbf{C}_t)$ its energy. We take

$$\boldsymbol{p_t} = \boldsymbol{p}(\mathbf{C}_t) \propto \frac{\mathbf{1}}{\widehat{\boldsymbol{\Omega}}(\boldsymbol{E_t})},$$

as its probability. The probability of acceptance of the trial state is then given by

$$\boldsymbol{p} \;=\; \mathbf{minimum}\left(\mathbf{1}, \frac{\boldsymbol{p_t}}{\boldsymbol{p_i}}\right) = \mathbf{minimum}\left(\mathbf{1}, \frac{\widehat{\boldsymbol{\Omega}}(\boldsymbol{E_i})}{\widehat{\boldsymbol{\Omega}}(\boldsymbol{E_t})}\right)$$

Note that if the trial micro state belongs to a lower entropy region it gets accepted with unit probability; however if it belongs to higher entropy region its acceptance probability is less than unity. Thus the algorithm pushes the Markov chain preferentially toward low entropy region. This preference cancels statistically exactly the natural tendency of randomly sampling of trial micro states from high entropy region. As a result the Markov chain shall have equal number of micro states in equal regions of energy. In other words the energy histogram of the visited micro states shall be flat.

Thus the Markov chain visits all regions of energy with equal ease. It does not see any energy barriers insurmountable or otherwise, that might be present in the system under investigation. This is a great advantage because there are indeed energy barriers that emerge close to first order phase transition and which are responsible for the super critical slowing of the dynamics. Also glassy systems have free energy profile with numerous ups and downs. Perhaps it is worthwhile investigating and inventing methods that that help obtain the density of states iteratively. May be we have to to abandon the comforts of Markov Chain methodology and violate detailed balance of the Metropolis rule to achieve this. This line of thoughts lead us naturally to non-Boltzmann Monte Carlo algorithms Also notice that if we manage to get the density of states then we get easy access to entropy and other thermal properties. It is indeed very difficult to estimate entropy and other thermal quantities by conventional Boltzmann Monte Carlo methods.

What is it that renders calculation of entropy a difficult task ? To answer this question we have to realize that the usefulness of the Monte Carlo methods considered upto now, is tied crucially to our ability to assign a numerical value of the property $\boldsymbol{O}$ to every micro state of the system. Consider estimating a property like entropy. We can not assign a numerical value for entropy to any single micro state of the system. All the micro states collectively own entropy. Hence thermal properties in general and entropy in particular are not easily accessible to conventional Monte Carlo methods based on Metropolis algorithm. Let us bring all the conventional Monte Carlo methods under one single umbrella and call them Boltzmann Monte Carlo methods.

For computing thermal properties we need to go beyond Boltzmann Monte Carlo methods. We need non-Boltzmann Monte Carlo methods.

That a non-Boltzmann sampling can provide a legitimate and perhaps superior alternative to Boltzmann methods has been recognized even during the very early days of Monte Carlo practice, see *e.g.* [8] and to these issues we turn our attention, below.

Torrie and Valleau [9] were, perhaps, the first to propose a non-Boltzmann algorithm to calculate the thermal properties. Their method called umbrella sampling has since undergone a series of metamorphoses. We have the multi-canonical Monte Carlo of Berg and Neuhaus [10], entropic sampling of Lee [11] and the algorithm of Wang and Landau [12].

We describe below the Wang-Landau algorithm.

3.1 Wang-Landau Algorithm

Wang and Landau [12] proposed an algorithm to estimate iteratively the density of states of the system. The algorithm is described below.

Define a function $\boldsymbol{g(E)}$ and set it to unity for all $\boldsymbol{E}$. Define also an histogram $\boldsymbol{H(E)}$ and set it to zero for all $\boldsymbol{E}$. Start with an arbitrary initial micro state $\mathbf{C_0}$. Let $\boldsymbol{E_0 = E(C_0)}$ be its energy. Update $\boldsymbol{g(E)}$ and $\boldsymbol{H(E)}$ as follows :

$$\boldsymbol{g(E_0) = g(E_0) \times \alpha; \quad H(E_0) = H(E_0) + 1.}$$

Here $\boldsymbol{\alpha}$ is the Wang-Landau factor and we take $\boldsymbol{\alpha = e^1 = 2.7183}$ in the first iteration. Construct a trial state and accept/reject it on the basis of Metropolis algorithm where we take

$$\boldsymbol{p(\mathrm{C}) \propto \frac{1}{g(E(\mathrm{C}))}.}$$

Every time we advance the chain, we update the functions $\boldsymbol{g(E)}$ and $\boldsymbol{H(E)}$. We start employing the updated $\boldsymbol{g}$ from the very next rejection/acceptance step. Carry out the simulation of the chain of micro states until the energy histogram of becomes flat over, at least, a small range of energy. This constitutes one Wang-Landau iteration.

Note that the function $\boldsymbol{g(E)}$ is updated at every step; the updated function is employed for decision making, from the the very next step. As a result the chain of micro states generated, is not Markovian. The probability of transition between two micro states at any stage in the chain depends on how many times the chain has visited these two micro states in its past. The transition from present to future depends on the entire past. Hence we shall refer to the sequence of micro states as simply a chain and not a Markov chain.

At the end of the first Wang-Landau iteration, change $\boldsymbol{\alpha}$ to $\boldsymbol{\sqrt{\alpha}}$. Reset $\boldsymbol{H(E)}$ to zero for all $\boldsymbol{E}$; but continue with $\boldsymbol{g(E)}$. Carry out the second Wang-Landau iteration. The histogram would spread out and would at the same time become flatter over a wider range of energy.

Upon further iterations the value of alpha will move closer and closer to unity. For example, after some twenty five iterations we shall have $\boldsymbol{\alpha = 1 + 3 \times 10^{-7}}$. The histogram of energy would become flat at least over the range of energy of interest after a few Wang-Landau iteration runs.

The flatter the histogram, closer would be $\boldsymbol{g(E)}$ to the true but unknown density of states $\boldsymbol{\widehat{\Omega}(E)}$. We take $\boldsymbol{g(E)}$ obtained at the end of the last iteration - the one which generates a reasonably flat energy histogram, as an estimate of $\boldsymbol{\widehat{\Omega}(E)}$, the true density of states.

We can define a suitable criteria for measuring the flatness of the histogram. For example we can consider the histogram to be flat if the smallest and largest entries do not differ from each other by say more than say ten percent. Depending upon the requirement of accuracy and the availability of computing resources, we can relax or tighten the flatness criterion.

There is no hard and fast rule about either the choice of the initial value of the Wang-Landau factor or its evolution from one iteration to the next. The choice of $\boldsymbol{\alpha = \alpha_0 = e^1}$ at the beginning of the first iteration and the square-root rule of evolution, were recommended by Wang and Landau[12]. In principle, $\boldsymbol{\alpha_0}$ can be any real number greater than unity and it should decrease,

preferably monotonically, to unity. Some authors, see *e.g.* [13, 14], have found it advantageous vary $\boldsymbol{\alpha}$ non-monotonically. The important point is any choice of variation of $\boldsymbol{\alpha}$ that flattens the histogram would serve the purpose. In a sense the histogram provides a diagnostic tool with which you can monitor whether you are doing things right or wrong. The flatness of the histogram tells you how close has the density of states converged to its true value.

The Wang-Landau algorithm estimates the density of states only upto a normalization constant. In other words the micro canonical entropy is estimated only upto an additive constant. This is quite adequate since we need to calculate only change in entropy rather than absolute entropy in almost all applications.

In principle we can stop here. Once we know the density of states then we can employ the machinery of thermodynamics and know everything else about the system.

Entropic Ensemble

Alternately, we can employ the converged density of states in a production run and generate a large ensemble of micro states. The sequence of micro states generated in the production run constitute a legitimate Markov chain, obeying detailed balance. However the invariant probabilities are unphysical : the probability of a micro state $\boldsymbol{\mathcal{C}}$ is inversely proportional the density of states at $\boldsymbol{E = E(\mathcal{C})}$. The Markov chain obeys detailed balance and hence convergence to the desired ensemble, though unphysical, is guaranteed.

Let us call the set of micro states generated in the production run as an an entropic ensemble or Wang-Landau ensemble. By employing un-weighting and re-weighting techniques[19] we can make, from the entropic ensemble, statically reliable estimates of physical quantities.

In what follows I shall show how to convert the entropic ensemble to a micro canonical ensemble and to a canonical ensemble.

Entropic → Micro Canonical Ensemble

Let $\boldsymbol{\{C_i \ : \ i = 1, 2, \cdots, M\}}$ denote a set of $\boldsymbol{M}$ micro states belonging to the entropic ensemble. We attach a statistical weight unity to each of these micro states : $\boldsymbol{\{W(C_i) = 1 \ : \ i = 1, 2, \cdots, M\}}$ These micro states have been sampled from a probability distribution $\boldsymbol{p(C_i) \propto 1/g(E(C_i))}$. Hence we first carry out un-weighting, see footnote (19) :

$$\boldsymbol{W(C_i) = \frac{W(C_i)}{1/g(E(C_i))} = W(C_i) \times g(E(C_i)).}$$

Note that the micro states of the entropic ensemble are not necessarily of the same energy. In fact the ensemble contains equal number of micro states in equal regions of energy - in other words the energy-histogram is flat. For a micro canonical ensemble all micro states are of the same energy and are equally probable. Hence we the re-weighting factor is $\boldsymbol{1 \times \delta(E - E(C_i))}$; the delta function ensures that we assemble only those micro statesEvery time we advance the chain, we update the functions $\boldsymbol{g(E)}$ and $\boldsymbol{H(E)}$. We start employing the updated $\boldsymbol{g}$ from the very Every time we advance the chain, we update the functions $\boldsymbol{g(E)}$ and $\boldsymbol{H(E)}$. We start employing the updated $\boldsymbol{g}$ from the very n n with the desired energy. Thus we have

$$\boldsymbol{W(C_i) = g(E(C_i)) \ \delta(E - E(C_i)).}$$

[19] See Appendix-F on Un-weighting and Re-weighting

Let $\boldsymbol{O}(\mathbf{C}_i)$ be the value of a property when the system is in micro state $\mathbf{C}_i$. The micro canonical ensemble average of $\boldsymbol{O}$ is given by,

$$\langle O\rangle_{\mu C}(E) \;=\; \underset{M\to\infty}{\text{Limit}} \frac{\sum_{i=1}^{M} O(\mathbf{C}_i)g(E(\mathbf{C}_i))\delta(E(\mathbf{C}_i)-E)}{\sum_{i=1}^{M} g(E(\mathbf{C}_i))\delta(E(\mathbf{C}_i)-E)}$$

In the above we have taken $\boldsymbol{E}$ as the energy of the isolated system described by the micro canonical ensemble.

Thus weighted averaging over micro states of given energy belonging to the *unphysical* entropic ensemble equals averaging over a *physical* micro canonical Every time we advance the chain, we update the functions $\boldsymbol{g(E)}$ and $\boldsymbol{H(E)}$. We start employing the updated $\boldsymbol{g}$ from the very Every time we advance the chain, we update the functions $\boldsymbol{g(E)}$ and $\boldsymbol{H(E)}$. We start employing the updated $\boldsymbol{g}$ from the very n nensemble modelling an isolated system.

Entropic → Canonical Ensemble

The un-weighting factor remains the same as the one derived for converting entropic ensemble to micro canonical ensemble. The re-weighting factor however is the Boltzmann weight. Thus

$$\boldsymbol{W(C_i) = g(E(\mathbf{C}_i)) \times \exp[-\beta E(\mathbf{C}_i)].}$$

All the micro states of the entropic ensemble contribute to the canonical ensemble average.

The canonical ensemble average of $\boldsymbol{O}$ is given by

$$\langle O\rangle_{C} \;=\; \underset{M\to\infty}{\text{Limit}} \frac{\sum_{i=1}^{M} O(\mathbf{C}_i)g(E(\mathbf{C}_i))\exp[-\beta E(\mathbf{C}_i)]}{\sum_{i=1}^{M} g(E(\mathbf{C}_i))\exp[-\beta E(\mathbf{C}_i)]}$$

Thus the weighted average over the *unphysical* entropic ensemble is equivalent to average over a *physical* canonical ensemble modelling a closed system.

From one single ensemble of micro states we can calculate averages over a large number of distinct canonical ensembles at different temperatures. This is a huge advantage especially for problems in which we need the properties on a fine grid of temperatures in the neighbourhood of a phase transition. Every time we advance the chain, we update the functions $\boldsymbol{g(E)}$ and $\boldsymbol{H(E)}$. We start employing the updated $\boldsymbol{g}$ from the very Every time we advance the chain, we update the functions $\boldsymbol{g(E)}$ and $\boldsymbol{H(E)}$. We start employing the updated $\boldsymbol{g}$ from the very n n

4 End Note

We have discussed about Metropolis algorithm to sample micro states from a given ensemble, physical or otherwise. If sampling is done from a physical ensemble we call it Boltzmann Monte Carlo. Boltzmann sampling has been eminently successful for estimating mechanical properties like energy. The reason is simple. A value for a mechanical property can be assigned to each micro state. All the micro states of the entropic ensemble contri However Boltzmann sampling is Every time we advance the chain, we update the functions $\boldsymbol{g(E)}$ and $\boldsymbol{H(E)}$. We start employing the updated $\boldsymbol{g}$ from theEvery time we advance the chain, we update the functions $\boldsymbol{g(E)}$ and $\boldsymbol{H(E)}$. We start employing the updated $\boldsymbol{g}$ from the very Every time we advance the chain, we update the functions $\boldsymbol{g(E)}$ and $\boldsymbol{H(E)}$. We start employing the updated $\boldsymbol{g}$ from the very n n very n quite clumsy when it comes to estimatEing thermal properties like entropy and free energies. The clumsiness owes its origin to the fact that a numerical value for entropy can not be assigned to any single micro state. All the

micro states, collectively, own entropy. Entropy is a property of an ensemble and not of any single micro state. This problem about estimating entropy was recognized even in the early days of Monte Carlo methods development and Torrie and Valleau[9] invented umbrella sampling which precisely and brilliantly addressed these issues. Umbrella sampling has since inspired and given rise to a whole host of non-Boltzmann methods; the latest to arrive on the Monte Carlo scene, is the method from Wang and Landau[12]. We have described the basic theory behind Wang-Landau algorithm and how to implement it on a practical problem.

The take home message is that non-Boltzmann Monte Carlo methods are as good as Boltzmann methods, if not more, for purpose of calculating mechanical properties. They also provide reliable estimates of thermal properties, not accessible to Boltzmann Monte Carlo methods.

We must quickly add that all is not cozy about the non-Boltzmann Monte Carlo methods. There are several issues and difficulties associated with the Wang-Landau algorithm in particular and non-Boltzmann Monte Carlo methods in general. A typical Monte Carlo aficionado, see *e.g.* [15], is not quite comfortable with the Wang-Landau algorithm since it does not obey detailed balance; in fact the chain generated is not even Markovian. What are the implications of these issues to the convergence of the density of states ? The convergence, we must say, has been unambiguously established numerically on several problems. But then do we have a proof that $\boldsymbol{g(E)}$ would converge to the true density of states $\boldsymbol{\Omega(E)}$?

Another problem is the algorithm performs miserably poorly if the Hamiltonian of the system under simulation is continuous. This problem has attracted the attention of several authors, se Every time we advance the chain, we update the functions $\boldsymbol{g(E)}$ and $\boldsymbol{H(E)}$. We start employing the updated $\boldsymbol{g}$ from the very ne *e.g.* [13, 14, 16, 17]. But to the best of our knowledge no saEtisfactory solution has yet emerged. All the suggestions that have been proposed so far, E including the method of frontier sampling seem ad-hoc at least in our perspective.

There are also issues about errors - both systematic and statistical - associated with the computed density of states. How does one translate the non-flatness of the energy histogram to error bars in the estimated density of states ? After all, a Monte Carlo practitioner often takes pride in in the fact that the method gives not only an estimate of a property but also of associated statistical error. But then, we do not know how to calculate Monte Carlo error bars in Wang-Landau simulation ? We hope these and other issues would get resolved soon and we shall have the Non-Boltzmann Monte Carlo methods on a strong theoretical and practical footing, sooner than later.

References

[1] N Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, *Equation of State Calculations by Fast CEomputing Machines*, Journal of Chemical Physics **21** 1087 (1953); see also G. Bhanot, *The Metropolis Algorithm*, Reports of Progress in Physics **51**, 429 (1988).

[2] K. P. N. Murthy and V. S. S. Sastry, *Markov chain Monte Carlo Methods in Statistical Physics*, PoS(SMPRI2005)018 (2005).

[3] W. Feller, *An Introduction to Probability Theory and Applications I and II*, John Wiley (1968)

[4] A. Papoulis, *Probabilty Theory, Random Variables, and Stochastic Processes* McGraw Hill (1965)

[5] K. P. N. Murthy, *Monte Carlo Methods in Statistical Physics*, Universities Press (2004)

[6] K. P. N. Murthy, *Monte Carlo : Basics*, Report ISRP - TD-3, Indian Society for Radiation Physics, Kalpakkam Chapter (2000); Esee arXiv:cond-mat/014215v1, 12 Apr. 2001

[7] E. J. McGrath, and D. C. Irving, *ETechnqiues for Efficient Monte Carlo Simulation* Volume III : Variance Reduction, Report SAI - 72 - 590 - LJ , Office of the Naval Research, Department of the Navy, Arlington, Virginia 22217 USA (March 1973)

[8] L. D. Fosdick, *Monte Carlo Computation on the Ising Lattice* in Methods of Computational Physics, Vol. 1, Editor B Adler, p.245 (1963).

[9] G. M. Torrie, and J. P. Valleau, *Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling*, Journal of Computational Physics **23** 187 (1977). E

[10] B. A. Berg, and T. Neuhaus, *Multicanonical ensemble: A new approach to simulate first-order phase transitions*, Physical Review Letters **68**, 9 (1992).E

[11] J. Lee, *New Monte Carlo algorithm: Entropic sampling*, Physical Review Letters **71**, 211 (1993); *Erratum*, **71**, 2353 (1993)

[12] F. Wang, and D. P. Landau, *Efficient, Multiple-Range Random Walk Algorithm to Calculate the Density of States,* Physical Review Letters **86**, 2050 (2001).

[13] D. Jayasri, V. S. S. Sastry, and K. P. N. Murthy, *Wang-Landau Monte Carlo simulation of isotropic-nematic transition in liquid crystals*, Physical Review E **72**, 36702 (2005).

[14] P. Poulin, F. Calvo, RE. Antoine, M. Broyer, P. Dugord, *Performances of Wang-Landau algorithms for continuous systems*, Physical Review E **73**, 56704 (2006).

[15] W. Janke, *Monte Carlo Simulations in Statistical Physics – From Basic Principles to Advanced Applications*, in *Order, Disorder and Criticality: Advanced Problems of Phase Transition Theory*, Vol. 3 (edited by Y. Holovatch), World Scientific (2012)pp. 93-166.

[16] B. J. Schulz, K Binder, M. Müller, and D. P. Landau, *Avoiding Boundary Effects in Wang-Landau Sampling* Physical Review E **67**, 67102 (2003).

[17] C. Zou, and R. N. Bhatt, *Understanding and Improving the Wang-LandauE Algorithm*, Physical Review E **72**, 25701(R) (2005).

Appendix - A

Internal Energy and the First Law of Thermodynamics:

In thermodynamics, internal energy is defined completely in terms of work done in adiabatic processes : Select a reference point $\boldsymbol{O}$ in the thermodynamic phase plane. Define a function $\boldsymbol{U}$ as follows. Assign an arbitrary value to $\boldsymbol{U(O)}$. Without loss of generality we take it as zero : $\boldsymbol{U(O) = 0}$. Consider a point $\boldsymbol{A}$. E Measure or calculate work done in an adiabatic process that takes the system from $\boldsymbol{O}$ to $\boldsymbol{A}$. Then define : $\boldsymbol{U(A) = U(O) + W_i^{ad.}}$ The superscript $\boldsymbol{ad.}$ tells that the process considered is adiabatic. Employ the convention : *work done on the system is positive and work done by the system is negative.* By considering adiabatic processes we can define $\boldsymbol{U}$ at all points on the phase plane. If there exists a point, say $\boldsymbol{B}$, whiEch is not ac Every time we advance the chain, we update the functions $\boldsymbol{g(E)}$ and $\boldsymbol{H(E)}$. We start employing the updated $\boldsymbol{g}$ from the very ncessible adiabatically from $\boldsymbol{O}$ then consider an adiabatic process that takes the system from $\boldsymbol{B}$ to $\boldsymbol{O}$ for purpose of defining $\boldsymbol{U}$: $\boldsymbol{U(B) = U(O) - W_{B\to O}^{ad.}}$. Then consider an arbitrary process from $\boldsymbol{C}$ to $\boldsymbol{D}$. Let $\boldsymbol{W}$ be the work done and $\boldsymbol{\Delta U = U(D) - U(C)}$. Then, $\boldsymbol{\Delta Q = \Delta U - W}$ is called heat and

this is a statement of the first law of thermodynamics. Heat is the difference between actual work and adiabatic work. Thus the first law of thermodynamics establishes the mechanical equivalence of heat. As an off-shoot, we get to define a thermodynamic property calledE the intEernal energy, denoted by the symbol $\boldsymbol{U}$

Appendix - B

Markov Chain

Consider a sequence of micro states visited by the system at discrete times starting from $\boldsymbol{C_0}$ at time $\mathbf{0}$. Let us denote the sequence by

$$\mathbf{C_0} \rightarrow \mathbf{C_1} \rightarrow \cdots \rightarrow \mathbf{C_{n-1}} \rightarrow \mathbf{C_n},$$

where the subscript denote the discrete time index. Our interest is to calculate the joint probability of the sequence. From Bayes' theorem we have,

$$\begin{aligned}\boldsymbol{P}(\ \mathbf{C_n}, \mathbf{C_{n-1}}, \cdots \mathbf{C_1}, \mathbf{C_0}) = \\ \boldsymbol{P}(\ \mathbf{C_n} \mid \mathbf{C_{n-1}}, \mathbf{C_{n-2}} \cdots \mathbf{C_1}, \mathbf{C_0}\ ) \times \boldsymbol{P}(\ \mathbf{C_{n-1}}, \mathbf{C_{n-2}} \cdots \mathbf{C_1}, \mathbf{C_0}\ ).\end{aligned}$$

If $\boldsymbol{P}\ (\ \mathbf{C_n} \mid \mathbf{C_{n-1}}, \mathbf{C_{n-2}}, \cdots \mathbf{C_1}, \mathbf{C_0}\ ) = \boldsymbol{P}\ (\ \mathbf{C_n} \mid \mathbf{C_{n-1}}\ )$, then $\mathbf{C_0} \rightarrow \mathbf{C_1} \rightarrow \cdots \mathbf{C_{n-1}} \rightarrow \mathbf{C_n}$ is a Markov chain : The future depends only on the present and not on the past. Thus, once the present is specified, the future is independent of the past.

Under Markovian condition, the expression for the joint probability of the chain of micro states, simplifies to

$$\begin{aligned}\boldsymbol{P}(\ \boldsymbol{C_n}, \boldsymbol{C_{n-1}}, \cdots \boldsymbol{C_1}, \boldsymbol{C_0}\ ) &= \boldsymbol{P}(\ \boldsymbol{C_n} \mid \boldsymbol{C_{n-1}}\ ) \times \boldsymbol{P}(\ \boldsymbol{C_{n-1}}, \boldsymbol{C_{n-2}} \cdots \boldsymbol{C_1}, \boldsymbol{C_0}\ ), \\ &= \boldsymbol{P}(\ \boldsymbol{C_n} \mid \boldsymbol{C_{n-1}}\ ) \times \boldsymbol{P}(\ \boldsymbol{C_{n-1}} \mid \boldsymbol{C_{n-2}}\ ) \times \boldsymbol{P}(\ \boldsymbol{C_{n-2}}, \boldsymbol{C_{n-3}} \cdots \boldsymbol{C_1}, \boldsymbol{C_0}\ ), \\ &= \cdots\cdots\cdots, \\ &= \boldsymbol{P}(\ \mathbf{C_n} \mid \mathbf{C_{n-1}}\ ) \times \boldsymbol{P}(\ \mathbf{C_{n-1}} \mid \mathbf{C_{n-2}}\ ) \times \cdots \times \boldsymbol{P}(\ \mathbf{C_1} \mid \mathbf{C_0}\ ) \times \boldsymbol{P}(\ \mathbf{C_0}\ ).\end{aligned}$$

Since we are interested in equilibrium properties we consider a sequence of states visited by an equilibrium system : The conditional probability, $\boldsymbol{P}(\ \mathbf{C_n} \mid \mathbf{C_{n-1}}\ )$ is independent of the time index. In other words $\boldsymbol{P}(\ \mathbf{C_n} = \mathbf{C_\mu} \mid \mathbf{C_{n-1}} = \mathbf{C_\nu}\ ) = \boldsymbol{W_{\mu,\nu}}$, and this quantity is independent of time. Once we know the transition probability matrix $\boldsymbol{W}$ and initial probabilities of all the micro states, we can calculate the probabi lity of any given Markov Chain. The transition probability matrix $\boldsymbol{W}$ is a square matrix of size $\widehat{\boldsymbol{\Omega}}$. We have

$$\mathbf{0} \leq \boldsymbol{W_{\mu,\nu}} \leq \mathbf{1}\ \forall\ \boldsymbol{\mu, \nu}$$

and

$$\sum_{\boldsymbol{\mu=1}}^{\widehat{\boldsymbol{\Omega}}} \boldsymbol{W_{\mu,\nu}} = \mathbf{1}\ \forall\ \boldsymbol{\nu}.$$

The elements of each column add to unity.

Appendix - C

Time Homogeneous Markov Chain

Let $\boldsymbol{P(\mathcal{C}_j, n)}$ be the probability for the system to be in micro state $\boldsymbol{\mathcal{C}_j}$ at discrete time $\boldsymbol{n}$. Let $\boldsymbol{W_{i,j}}$ denote the probability for transition from micro state $\boldsymbol{\mathcal{C}_j}$ to micro state $\boldsymbol{\mathcal{C}_i}$, in one time step. We consider time homogeneous Markov chain for which the transition probabilities $\boldsymbol{\{W_{i,j} : i, j = 1, 2, \cdots \widehat{\Omega}\}}$ are independent of time. We have $\boldsymbol{W_{i,j} = P\ (\ \mathcal{C}_i \mid \mathcal{C}_j\ )}$, the conditional probability that the system is in micro state $\boldsymbol{\mathcal{C}_i}$ at any instant of time given it was in micro state $\boldsymbol{\mathcal{C}_j}$ at the previous instant of time.

Master Equation : The probabilities obey the Master equation,

$$\boldsymbol{P(\mathcal{C}_i; n+1)} \ = \ \sum_{j\ :\ j \neq i} \boldsymbol{P(\mathcal{C}_j, n) W_{i,j} + P(\mathcal{C}_i, N)\ W_{i,i}}$$

We have $\sum_i \boldsymbol{W_{i,j} = 1\ \forall\ j}$. Therefore, $\boldsymbol{W_{i,i} = 1 - \sum_{j\ :\ j \neq i} W_{j,i}}$. We can write the Master equation as

$$\begin{aligned} \boldsymbol{P(\mathcal{C}_i; n+1)} \ &= \ \sum_{j \neq i} \boldsymbol{P(\mathcal{C}_j, n)\ W_{i,j} +\ P(\mathcal{C}_i, n)} \left(1 - \sum_{j\ :\ j \neq i} \boldsymbol{W_{j.i}} \right) \\ &= \ \boldsymbol{P(\mathcal{C}_i, n)} + \sum_{j \neq i} \boldsymbol{[P(\mathcal{C}_j, n)\ W_{i,j} - P(\mathcal{C}_i, n) W_{j,i}]} \end{aligned}$$

Balance Condition : When the system equilibrates we have

$$\boldsymbol{P(\mathcal{C}_i, n+1) = P(\mathcal{C}_i, n) = p(\mathcal{C}E_i)\ \forall\ i.}$$

Therefore we have

$$\sum_j \boldsymbol{\left[\ p(\mathcal{C}_j)\ \times\ W_{i,j}\ -\ p(\mathcal{C}_i)\ \times\ W_{j,i}\ \right] = 0.}$$

This is called the balance condition which ensures that the Markov chain eventually equilibrates.

Detailed Balance

Look at the balance condition given above. We can make a stricter demand that each term in the sum be zero. Then we get the detailed balance condition :

$$\boldsymbol{p(\mathcal{C}_j)\ \times\ W_{i,j}\ =\ p(\mathcal{C}_i)\ \times\ W_{j,i} \quad \forall\ i, j = 1, 2, \cdots, \widehat{\Omega}.}$$

It is quite easy to show that the Metropolis rejection algorithm obeys detailed balance condition.

Appendix - D

Time Symmetry

By observing an equilibrium system we can not tell which direction time flows. Both directions are equally probable and equally unverifiable. Consider a Markov chain of micro states visited by an

equilibrium system : $\mathbf{C_0} \rightarrow \mathbf{C_1} \rightarrow \cdots \mathbf{C}_n \rightarrow \mathbf{C}_{n+1} \rightarrow \cdots \mathbf{C}_M$. The transition probabilities are given by $\boldsymbol{W_{i,j}} = \boldsymbol{P}(\mathbf{C}_n = \boldsymbol{\mathcal{C}_i}|\mathbf{C}_{n-1} = \boldsymbol{\mathcal{C}_j})$

At discrete time $\boldsymbol{M}$ let us reverse the Markov chain and get $\mathbf{C}_M \rightarrow \mathbf{C}_{M-1} \rightarrow \cdots \mathbf{C}_{n+1} \rightarrow \mathbf{C}_n \rightarrow \cdots \mathbf{C_1} \rightarrow \mathbf{C_0}$. A little thought will tell you the above is also a Markov chain : for, the future depends only on the present and not on the past for a Markov chain, Hence once the present is specified the future is independent of the past. Past is independent of the future which renders the time reversed chain, Markovian.

Let us denote the transition probability matrix of the time reversed chain by the symbol $\boldsymbol{W^R}$. We have

$$\begin{aligned} \boldsymbol{W^R_{i,j}} = \boldsymbol{P}(\mathbf{C}_n = \boldsymbol{\mathcal{C}_i}|\mathbf{C}_{n+1} = \boldsymbol{\mathcal{C}_j}) &= \frac{\boldsymbol{P}(\mathbf{C}_n = \boldsymbol{\mathcal{C}_i},\ \mathbf{C}_{n+1} = \boldsymbol{\mathcal{C}_j})}{\boldsymbol{p}(\boldsymbol{\mathcal{C}_j})} \\ &= \frac{\boldsymbol{P}(\mathbf{C}_{n+1} = \boldsymbol{\mathcal{C}_j}|\mathbf{C}_n = \boldsymbol{\mathcal{C}_i})\ \boldsymbol{p}(\boldsymbol{\mathcal{C}_i})}{\boldsymbol{p}(\boldsymbol{\mathcal{C}_j})} \\ &= \frac{\boldsymbol{W_{j,i}}\ \boldsymbol{p}(\boldsymbol{\mathcal{C}_i})}{\boldsymbol{p}(\boldsymbol{\mathcal{C}_j})} \end{aligned}$$

The condition for reversibility is $\boldsymbol{W^R_{i,j}} = \boldsymbol{W_{i,nj}}$: The transition probability matrix should be the sane for both Markov chains - the time forward and the time reversed. Hence on the left hand side of the above equation replace $\boldsymbol{W^R_{i,j}}$ by $\boldsymbol{W_{i,j}}$ and Every time we advance the chain, we update the functions $\boldsymbol{g(E)}$ and $\boldsymbol{H(E)}$. We start employing the updated $\boldsymbol{g}$ from the very n reorganize. Then the condition for reversibility reads as,

$$\boldsymbol{W_{i,j}}\ \boldsymbol{p}(\boldsymbol{\mathcal{C}_j}) \ = \ \boldsymbol{W_{j.i}}\ \boldsymbol{p}(\boldsymbol{\mathcal{C}_i}).$$

We recognize this as the detailed balance condition, see Appendix - C on Time Homogeneous Markov Chain. Thus a Markov chain of micro states of an equilibrium system obeys detailed balance condition and hence is reversible;

Apendix - E

Central Limit Theorem (CLT)

Let $\boldsymbol{Y} = \boldsymbol{X_1} + \boldsymbol{X_2} + \cdots \boldsymbol{X_M}$ denote the sum of $\boldsymbol{N}$ independent random variables each of finite mean and variance. The central limit theorem says that $\boldsymbol{Y}$ is a Gaussian in the limit of $\boldsymbol{M} \rightarrow \boldsymbol{\infty}$ and the variance of the Gaussian distribution is proportional to $\boldsymbol{M}$. The Central Limit Theorem is a glorious culmination of a series of studies starting from the Chebyshev inequality, see *e.g.* [3, 4] followed by various laws of large numbers, see any standard text book, see *e.g.* [3, 4] on Probability theory to know about these issues. Chebyshev inequality tells us the probability that a realization of a random variable would deviate from its mean beyond $\boldsymbol{k\sigma}$, is less than $\boldsymbol{1/k^2}$ for $\boldsymbol{k \geq 1}$, where $\boldsymbol{\sigma}$ is the standard deviation of the random variable.

Apendix - F

Un-Weighting and Re-Weighting

Let me explain un-weighting and re-weighting in a simple manner[5, 6]. Let $\boldsymbol{x}$ be a random variable and $\boldsymbol{f(x)}$ its probability density. Let $\boldsymbol{h(x)}$ be some function of $\boldsymbol{x}$. The $\boldsymbol{f}$-ensemble average of $\boldsymbol{h}$ is

ISSN(online):2320-9836
ISSN(Print):2320-9828

formally expressed as,

$$\langle \boldsymbol{h} \rangle_{\boldsymbol{f}} = \int_{-\infty}^{+\infty} \boldsymbol{dx}\ \boldsymbol{h(x)}\ \boldsymbol{f(x)}.$$

Let $\boldsymbol{g(x)}$ be a density function. We call it importance density function. Let us generate an ensemble,

$$\boldsymbol{\Omega_g = \{x_i\ :\ i = 1, 2,\ \cdots M\}}$$

by random sampling from $\boldsymbol{g(x)}$.

Our aim is to make an estimate of $\langle \boldsymbol{h} \rangle_{\boldsymbol{f}}$ employing the ensemble $\boldsymbol{\Omega_g}$.

Consider the following.

$$\langle \boldsymbol{h} \rangle_{\boldsymbol{f}} \ = \ \int_{-\infty}^{+\infty} \boldsymbol{dx}\ \boldsymbol{h(x) f(x)} \ = \int_{-\infty}^{+\infty} \boldsymbol{dx}\ \boldsymbol{h(x)} \frac{\boldsymbol{f(x)}}{\boldsymbol{g(x)}} \boldsymbol{g(x)} \ = \langle \ \boldsymbol{h(1/g) f}\ \rangle_{\boldsymbol{g}}$$

The above is an exact result. The left hand side is an $\boldsymbol{f}$ ensemble average of $\boldsymbol{h}$. The right hand side is a $\boldsymbol{g}$ ensemble average of $\boldsymbol{h}$ unweighted by $\boldsymbol{1/g}$ and re-weighted by $\boldsymbol{f}$.

The implementation goes as follows.

$$\langle \boldsymbol{h} \rangle_{\boldsymbol{f}} = \underset{M \to \infty}{\textbf{Limit}} \frac{\boldsymbol{1}}{\boldsymbol{M}} \sum_{\boldsymbol{i=1}}^{\boldsymbol{M}} \boldsymbol{h(x_i)} \ \times \frac{\boldsymbol{1}}{\boldsymbol{g(x_i)}} \ \times \ \boldsymbol{f(x_i)}.$$

The Publication fee is defrayed by Indian Society for Education and Environment (www.iseeadyar.org)
Cite this article as:
T. Pramananda Perumal, K. P. N. Murthy. Metropolis and Wang-Landau Algorithms. *Indian Journal of Economics and Development*, Vol 6(2), February 2018.