© Indian Academy of Sciences

CrossMark

Non-Boltzmann ensembles and Landau free energy

M. SUMAN KALYAN[1] and K. P. N. MURTHY[2,*]

[1]Department of Physical Sciences, Indian Institute of Science Education and Research Mohali, Sector 81, Sahibzada Ajit Singh Nagar, Manauli 140 306, India
[2]Department of Physics, School of Physical Sciences, Central University of Rajasthan, Bandarsindri, N. H. 8 (Jaipur-Ajmer Highway), Kishangargh 305 817, Ajmer, India
*Corresponding author. E-mail: kpn.murthy@curaj.ac.in

Abstract. Boltzmann sampling based on the Metropolis algorithm has been extensively used for simulating a canonical ensemble. An estimate of a mechanical property, like energy, of an equilibrium system, can be made by averaging over a large number of microstates generated by Boltzmann Monte Carlo methods. However, a thermal property like entropy is not easily accessible to these methods. The reason is simple. We can assign a numerical value for energy to each microstate. But we cannot make such an assignment for entropy. Entropy is not a property associated with any single microstate. It is a collective property of all the microstates. Towards calculating entropy and other thermal properties, a non-Boltzmann Monte Carlo technique called Umbrella sampling was proposed in the mid-seventies (of the last century). Umbrella sampling has since undergone several metamorphoses and we have now, multicanonical Monte Carlo, entropic sampling, flat histogram methods, Wang–Landau algorithm *etc.* This class of methods generates non-Boltzmann ensembles which are unphysical. However, physical quantities can be calculated by un-weighting the microstates of the entropic ensemble, followed by re-weighting to the desired physical ensemble.

In this review we shall tell you of the Metropolis algorithm for estimating the mechanical properties and of the Wang–Landau algorithm for estimating both mechanical and thermal properties of an equilibrium system. We shall demonstrate the utility of non-Boltzmann Monte Carlo methods by calculating Landau free energy in a model system consisting of q-state Potts spins on a two-dimensional square lattice. The model exhibits a first-order phase transition for $q > 4$ and a second-order phase transition for $q \leq 4$. We report results on the Potts spin model for $q = 8$ (first-order phase transition) and for $q = 2$ (second-order phase transition). We also present the results on a more realistic problem of temperature-induced unbinding or denaturation of a hairpin DNA. The transition from a closed phase to an open phase is found to be first order.

We shall attempt to make this review as pedagogical and self-contained as possible.

Keywords. Monte Carlo methods; entropy; free energy; ensembles; random sampling; Metropolis algorithm; entropic sampling.

PACS Nos 05.10.Ln; 87.10.Rt; 87.16.a; 87.16.A; 87.16Ac

1. Some preliminaries

Statistical mechanics helps us go from the microworld of atoms and molecules obeying laws[1] of classical and quantum mechanics to the macroworld of thermodynamics[2] describing matter in bulk. The micro–macro synthesis proceeds, very generally, along the following lines.

First, we identify a random variable that corresponds to a thermodynamic property. The average of the random variable over a suitable and well-defined statistical ensemble[3] gives the value of the thermodynamic property.

---

[3]*e.g.* Gibbs' ensembles: microcanonical for an isolated system; canonical for a closed system and grand canonical for an open system.

Maxwell's ensemble:
The notion of an ensemble was first proposed by James Clerk Maxwell (1831–1879). A Maxwell's ensemble is a set, whose elements are drawn from the microstates of the system under considerations. A microstate occurs in the ensemble several times. Number of times it occurs divided by the size of the ensemble equals its probability. Thus, an ensemble contains, not only information about the microstates of the system, but also about their probabilities.

---

[1]Time-reversal invariant.
[2]With an arrow of time given by the direction of increasing entropy.

As an example, consider internal energy[4]. This property is usually denoted by the symbol U. Corresponding to this property, we have, in statistical mechanics, a random variable E which is the sum of kinetic energy and the interaction energy of the atoms and molecules of the macroscopic object. A numerical

value for E can be assigned to each microstate[5] of the macroscopic system. The value of E will fluctuate when the equilibrium system goes from one microstate to another. These fluctuations, given by $\sigma_E^2 = \langle E^2 \rangle - \langle E \rangle^2$ also constitute an integral part of an equilibrium description. The symbol $\langle (\cdot) \rangle$ denotes averaging of the property $(\cdot)$ over the chosen ensemble. We have $\langle E \rangle = U$, and $\sigma_E^2 = k_B T^2 C_V$; the latter is called the fluctuation dissipation theorem[6]. k_B is the Boltzmann constant[7] and C_V is the heat capacity of the system.

1.1 *Energy*

The computation of average energy is now straightforward: Generate a canonical ensemble employing, for example, the Monte Carlo method based on the Metropolis algorithm; a simple arithmetic average of 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 microstate of the macroscopic system.

1.2 *Entropy*

How does one calculate entropy?

We cannot assign a numerical value for entropy to any single microstate. Entropy is a property that

---

footnote continued

Gibbs' ensemble:

Imagine now, a collection of a large number of identical mental copies of the macroscopic system under consideration. They constitute a **Gibbs' ensemble**. Each member of the Gibbs' ensemble shall be in one microstate or the other. Different members can be in different microstates; but all of them have the same macroscopic properties. This is what we mean when we say they are identical mental copies of the system. When the number of elements in the Gibbs' ensemble is large then the number of systems in a given microstate divided by the size of the ensemble will give the probability of the microstate.

Given the microstates and their probabilities, we can construct an ensemble employing for *e.g.* Monte Carlo methods. We can also do the reverse. Given an ensemble, we can calculate the probabilities of the microstates of the macroscopic system to which the ensembles belong. This is what we do when we derive Boltzmann weight for the microstates employing the method of the most probable distribution where we construct a canonical ensemble; see any standard text book on statistical mechanics *e.g.* [1].

[4] **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 O in the thermodynamic phase plane. Define a function U as follows. Assign an arbitrary value to $U(O)$. Consider a point A. Measure or calculate work done in an adiabatic process that takes the system from O to A. Then define:

$$U(A) = U(O) + W_{O \to A}^{\text{adi.}}$$

The superscript 'adi.' shows 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 U at all points on the phase plane. If there exists a point, say B, which is not accessible adiabatically from O then consider an adiabatic process that takes the system from B to O for the purpose of defining U:

$$U(B) = U(O) - W_{B \to O}^{\text{adi.}}.$$

Then consider an arbitrary process from C to D. Let W be the work done and $\Delta U = U(D) - U(C)$. Then,

$$\Delta Q = \Delta U - W$$

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 of the first law of thermodynamics we get to define a thermodynamic property called the internal energy, denoted by the symbol U, of a thermodynamic system.

---

[5] For example three positions (q_1, q_2, q_3) and three momenta (p_1, p_2, p_3) are required to specify a single-point particle. For N particles, we need a string of $6N$ numbers and this string denotes a microstate of the macroscopic system of N particles.

$$E = \frac{1}{2m} \sum_{i=1}^{3N} p_i^2 + V(q_1, q_2, \ldots, q_{3N}).$$

The energy consists of the kinetic energy and potential energy. Note that energy is defined for each microstate. For a macroscopic system of say N Ising spins, we have 2^N microstates since each Ising spin can be in either 'up' ($S_i = +1$) state or 'down' ($S_i = -1$) state.

$$E = -J \sum_{\langle i,j \rangle} S_i S_j,$$

where S_i is the spin at the lattice site i and $J > 0$ measures the strength of spin–spin interaction. Spins on the nearest neighbour lattice sites interact. The sum runs over all pairs of nearest-neighbour spins.

[6] Fluctuation dissipation theorem relates equilibrium fluctuations to the response of the system to small perturbation.

[7] $k_B = 1.38064852 \times 10^{-23}$ joules/kelvin. The Boltzmann constant helps us convert energy measured in units of kelvin to energy in units of joule.

belongs collectively to all the microstates. While energy is a private property (of each microstate), entropy is a social or a pubic property (see below).

Let $\Omega = \{\mathcal{X}_\nu : \nu = 1, 2, \ldots, \widehat{\Omega}\}$ denote the set of microstates of an equilibrium system; the microstates are discrete, distinct and finite in number. $\{p(\mathcal{X}_\nu) : \nu = 1, 2, \ldots, \widehat{\Omega}\}$ are their probabilities. We use 'script $\mathcal{X}$' to denote microstates of the system and 'Roman X' to denote microstates of an ensemble or of a Monte Carlo sample or of a Markov chain. The Boltzmann–Gibbs–Shannon entropy is given by

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

In the above, k_B is the Boltzmann constant.

2. Markov chain Monte Carlo simulation

Our aim is to simulate microscopic processes occurring in an equilibrium system and assemble a large number of microstates consistent with the given probabilities. To this end, we start with an arbitrary initial microstate $X_0 (\in \Omega)$; then, we generate a Markov chain[8], employing the Metropolis rejection algorithm [2].

---

[8] **Markov chain and Markov matrices**: Consider a sequence of microstates visited by the system at discrete times starting from X_0 at time 0. Let us denote the sequence by $X_0 \to X_1 \to \cdots \to X_{n-1} \to X_n$, where the subscripts denote the discrete-time index. Our interest is to calculate the joint probability of the sequence. From Bayes' theorem we have,

$$P(X_n, X_{n-1}, \ldots, X_0) = P(X_n \mid X_{n-1}, X_{n-2}, \ldots, X_0)$$
$$\times P(X_{n-1}, X_{n-2}, \ldots, X_0).$$

If $P(X_n \mid X_{n-1}, X_{n-2}, \cdots, X_0) = P(X_n \mid X_{n-1})$, then $X_0 \to X_1 \to \cdots X_{n-1} \to X_n$ 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 the Markovian condition, the expression for the joint probability of the chain of microstates simplifies to

$$P(X_n, X_{n-1}, \ldots, X_0) = P(X_n \mid X_{n-1})$$
$$\times P(X_{n-1}, X_{n-2}, \ldots, X_0)$$
$$= P(X_n \mid X_{n-1}) \times P(X_{n-1} \mid X_{n-2})$$
$$\times P(X_{n-2}, X_{n-3}, \ldots, X_0)$$
$$= \cdots \cdots \cdots$$
$$= \cdots \cdots \cdots$$
$$= P(X_n \mid X_{n-1}) \times P(X_{n-1} \mid X_{n-2}) \times \cdots$$
$$\cdots \times P(X_1 \mid X_0) \times P(X_0).$$

Since we are interested in equilibrium properties we demand the conditional probability, $P(X_n \mid X_{n-1})$ to be independent of the time index.

In other words

$$P(X_n = \mathcal{X}_\mu \mid X_{n-1} = \mathcal{X}_\nu) = W_{\mu,\nu},$$

The Markov chain is given by

$$X_0 (\in \Omega) \to X_1 (\in \Omega) \to X_2 (\in \Omega) \to \cdots$$
$$\cdots \to X_i (\in \Omega) \to X_{i+1} (\in \Omega) \to \cdots$$

In the above, the subscript is the time index. Each member of the chain is drawn from the set Ω of microstates.

2.1 *Metropolis rejection algorithm*

Let us say we have simulated the Markov chain up to $X_i \in \Omega$ starting from $X_0 \in \Omega$. Thus the current microstate is X_i. Let $p_i = p(X_i)$ denotes its probability. We make a small random change in the current microstate and construct a trial microstate[9] $X_t \in \Omega$. Let $p_t = p(X_t)$ denotes its probability. Calculate $p = \min(1, \, p_t/p_i)$. Then, the next microstate in the Markov chain is given by

$$X_{i+1} = \begin{cases} X_t & \text{with probability} \quad p \\ X_i & \text{with probability} \quad 1 - p \end{cases}$$

The implementation of the above is as follows:

- Generate a random number[10] uniformly distributed between zero and unity. Denote it by the symbol ξ.

- If $\xi \leq p$, then accept the trial state and advance the Markov chain to $X_{i+1} = X_t$.

---

and this quantity is independent of time. We call it time homogeneous Markov chain. Once we know the transition probability matrix W and initial probabilities of all the microstates, we can calculate the probability of any given Markov Chain. The transition probability matrix W is a square matrix of size $\widehat{\Omega}$. We have

$$0 \leq W_{\mu,\nu} \leq 1 \;\; \forall \;\; \mu, \nu \quad \text{and} \quad \sum_{\mu=1}^{\widehat{\Omega}} W_{\mu,\nu} = 1 \; \forall \; \nu.$$

W is called the Markov matrix or stochastic matrix. Its elements are all between zero and unity; the elements of each column add to unity. Besides, if the elements of each row also add to unity, then we have a doubly stochastic matrix.

[9] For example if we are simulating an Ising spin system, select randomly an Ising spin from the current spin configuration (microstate) 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 position coordinates and three-momentum coordinates by small random amounts to construct a trial microstate.

[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.

- If not, reject the trial state and advance the Markov chain to $X_{i+1} = X_i$.

- Repeat the process on X_{i+1} to get X_{i+2}; and so on.

Generate a long Markov chain. The asymptotic part of the chain shall contain microstates belonging to the ensemble characterized by the probabilities $\{p(X_\nu): \nu = 1, 2, \ldots\}$.

2.2 *Important properties of Metropolis algorithm*

The Metropolis algorithm demands only a knowledge of the ratio of probabilities of two microstates belonging to Ω. We should know this ratio for all possible pairs of microstates of Ω. This implies that we need to know $\{p(\mathcal{X}_\nu) : \nu = 1, 2, \ldots, \widehat{\Omega}\}$ 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 $\exp[-\beta E(\mathcal{X})]$ of each microstate; we need not have any knowledge what so ever of the canonical partition function.

The Metropolis algorithm obeys the balance condition[11]. The balance condition represents that the

---

[11] We consider a time-homogeneous Markov chain, see footnote (8). Let $P(\mathcal{X}_j, n)$ be the probability for the system to be in microstate $\mathcal{X}_j$ at discrete time n. Let $W_{i,j}$ denotes the probability for transition from the microstate $\mathcal{X}_j$ to the microstate $\mathcal{X}_i$ in a one-time step. We have

$$W_{i,j} = P(\mathcal{X}_i \mid \mathcal{X}_j),$$

the conditional probability that the system is in the microstate $\mathcal{X}_i$ at any instant of time given it was in the microstate $\mathcal{X}_j$ at the previous instant of time. The probabilities obey the equation given below.

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

We have $\sum_i W_{i,j} = 1 \ \forall \ j$. Therefore,

$$W_{i,i} = 1 - \sum_{j:\, j\neq i} W_{j,i}.$$

We can write,

$$P(\mathcal{X}_i; n+1) = \sum_{j\neq i} P(\mathcal{X}_j, n)\, W_{i,j} + P(\mathcal{X}_i, n)\left(1 - \sum_{j:\, j\neq i} W_{j,i}\right)$$

$$= P(\mathcal{X}_i, n) + \sum_{j\neq i} [P(\mathcal{X}_j, n)\, W_{i,j} - P(\mathcal{X}_i, n) W_{j,i}]$$

$$= P(\mathcal{X}_i, n) + \sum_{j=1}^{\widehat{\Omega}} [P(\mathcal{X}_j, n)\, W_{i,j} - P(\mathcal{X}_i, n) W_{j,i}] \ \forall \ i.$$

Balance condition: When the system equilibrates we have $P(\mathcal{X}_i, n+1) = P(\mathcal{X}_i, n) = \pi_i \ \forall \ i$. Therefore we have

$$\sum_j [\pi_j \times W_{i,j} - \pi_i \times W_{j,i}] = 0.$$

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

Markov chain shall converge, definitely, to an invariant probability distribution.

The Metropolis algorithm obeys a stricter condition, called detailed balance[12]. The consequences of this are two-fold.

1. The detailed balance ensures that the Markov chain generated converges to an equilibrium ensemble consistent with the given probability weights of the microstates[13]. Thus by choosing the probability weights properly we can *a priori* select the nature of the asymptotic ensemble of the Markov chain states.

2. The detailed balance ensures that the Markov chain generated is time reversible[14]; hence it is suited

---

[12] **Detailed balance**:
Look at the balance condition given toward the end of footnote 11 as a sum over j for each i. We can make a stricter demand that each term in the sum be zero. Then we get the detailed balance condition:

$$\pi_j \times W_{i,j} = \pi_i \times W_{j,i} \quad \forall \ i, j = 1, 2, \ldots, \widehat{\Omega}.$$

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

[13] Boltzmann weights, $\exp[-\beta E(\mathcal{X}_\nu)]$ for canonical ensemble; Gibbs weights, $\exp[-\beta(E(\mathcal{X}_\nu) - \mu N(\mathcal{X}_\nu))]$ for grand canonical ensemble; weights given by inverse of the density of states $1/\widehat{\Omega}(E(\mathcal{X}_\nu))$ for entropic ensemble, etc.

[14] By observing an equilibrium system we cannot tell which direction time flows. Both directions are equally probable and equally unverifiable. Consider a Markov chain of microstates visited by an equilibrium system: $X_0 \to X_1 \to \cdots X_n \to X_{n+1} \to \cdots X_M$. The transition probabilities are given by

$$W_{i,j} = P(X_n = \mathcal{X}_i \mid X_{n-1} = \mathcal{X}_j).$$

At discrete time M let us reverse the Markov chain and get

$$X_M \to X_{M-1} \to \cdots X_{n+1} \to X_n \to \cdots X_1 \to X_0.$$

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 W^R. We have

$$W^R_{i,j} = P(X_n = \mathcal{X}_i | X_{n+1} = \mathcal{X}_j) = \frac{P(X_n = \mathcal{X}_i,\ X_{n+1} = \mathcal{X}_j)}{p(\mathcal{X}_j)}$$

$$= \frac{P(X_{n+1} = \mathcal{X}_j | X_n = \mathcal{X}_i)\ p(\mathcal{X}_i)}{p(\mathcal{X}_j)}$$

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

The condition for reversibility is $W^R_{i,j} = W_{i,j}$: The transition probability matrix should be the same for both Markov chains – the time

for describing an equilibrium system; for, no matter what kind of observations you make on an equilibrium system, you cannot tell which way time moves. Equilibrium is a time-reversal invariant state. The detailed balance captures this subtle property.

2.3 *Estimation of averages and statistical errors*

Generate a Markov chain until it equilibrates[15]. Continue the Markov chain and collect a reasonably large number of microstates $\{X_i : i = 1, 2, \ldots, M\}$ from the equilibrated Markov chain. Let O be a property of interest and $O(X)$ its value when the system is in microstate X. Then the Monte Carlo estimate of the property O is given by[16]

$$\overline{O}_M = \frac{1}{M} \sum_{i=1}^{M} O\,(X_i); \quad \underset{M\to\infty}{\text{Limit}}\ \overline{O}_M = \langle O \rangle.$$

A little thought will tell you that the quantity $\overline{O}_M$ 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[17] (CLT) tells that the quantity $\overline{O}_M$ is a Gaussian random variable when M is large.

---

footnote continued

forward and the time reversed. Hence on the left hand side of the above equation replace $W_{i,j}^R$ with $W_{i,j}$ and reorganize. The condition for reversibility reads as, $W_{i,j}\ p(X_j) = W_{j,i}\ p(X_i)$. We immediately recognize this as the detailed balance condition, see footnote (12). Thus a Markov chain of microstates of an equilibrium system obeys the detailed balance condition and hence is reversible;

[15]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.

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

$$\langle O \rangle = \frac{1}{Q} \sum_{\nu=1}^{\widehat{\Omega}} O\,(X_\nu)\ \exp[-\beta E(X_\nu)]; \quad Q = \sum_{\nu=1}^{\widehat{\Omega}} \exp[-\beta E(X_\nu)].$$

[17]**Central limit theorem (CLT):** Let $X_1, X_2, \ldots, X_M$ be identically distributed independent random variables with finite mean, μ and finite variance, σ^2. Let

$$Y = \frac{X_1 + X_2 + \cdots + X_M}{M}.$$

The CLT shows that Y is a Gaussian with mean μ and variance σ^2/M when $M \to \infty$. The CLT is a glorious culmination of a series of studies starting with the Chebyshev inequality, see e.g. [3, 4]: A single number randomly sampled from a distribution, with finite mean μ, and finite variance σ^2, can fall outside the interval $\mu \pm k\sigma$ with a probability not more than $1/k^2$. Then came several laws of large numbers and these led eventually to the CLT, see any standard text book, e.g. [3, 4] on probability theory and stochastic processes to know more on these issues.

The average of the Gaussian is $\langle O \rangle$ and its variance is σ^2/M, where $\sigma^2 = \langle O^2 \rangle - \langle O \rangle^2$. A possible statement of the statistical error associated with the Monte Carlo estimate $\overline{O}_M$ is obtained from the following considerations.

Take a Gaussian random variable with mean ζ and standard deviation Σ. The area under the Gaussian[18] between $\zeta - \Sigma$ and $\zeta + \Sigma$ is 0.682695. This means that with 68.27% confidence, you can say that a randomly sampled number from the Gaussian shall lie between $\zeta - \Sigma$ and $\zeta + \Sigma$. 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,

$$\overline{O}_M^2 = \frac{1}{M} \sum_{i=1}^{M} O^2(X_i); \quad \underset{M\to\infty}{\text{Limit}}\ \overline{O}_M^2 = \langle O^2 \rangle.$$

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

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

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

$$\langle O \rangle - \frac{\sigma_M}{\sqrt{M}} \text{ and } \langle O \rangle + \frac{\sigma_M}{\sqrt{M}}.$$

The statistical error decreases with an increase of M. This is indeed a desirable property. This tells us, at least in principle, that we will get things right if 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 than the experimental error bar.

However, note that the statistical error decreases painfully slowly with an increase of the sample size. A 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 M.

We need variance reduction devices that significantly reduce the fluctuations without affecting the averages.

---

[18]

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

Importance sampling is a variance reduction device. It helps us sample microstates from important regions of the sample space e.g. microstates with high-Boltzmann weights. Note that a randomly selected microstate would be, most likely, of high energy[19], hence of low-Boltzmann weight; its contribution to the partition sum would be negligible. In fact the Metropolis algorithm is an importance sampling device. We are not going to discuss here about the importance sampling or other variance reduction techniques; those interested can consult [5–7].

Instead, in what follows, we shall investigate the nature of the invariant distribution of the Markov chain of microstates whose probabilities are inversely proportional to the density of states: microstates of the high-entropy region have low probabilities; and those of the low-entropy region have high probabilities. This kind of prescription does not describe any physical system. Nevertheless, constructing a Markov chain with these probabilities for the microstates has certain advantages and this will become clear in the sequel.

3. Markov chain with flat energy histogram

Consider a system with microstates

$$\Omega = \{\mathcal{X}_\nu : \nu = 1, 2, \ldots, M\}.$$

Let $\widehat{\Omega}(E)$ denotes its density of states. For the purpose of illustration we assume that the density of states is known. Let $\mathcal{X}_\mu \in \Omega$ and $E_\mu = E(\mathcal{X}_\mu)$. We prescribe

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

We emphasize two points, at the risk of being repetitive.

- We do not know the density of states beforehand[20].

- There is no physical system for which the probability of a microstate is inversely proportional to the density of states[21].

---

[19] Entropy increases with energy.

[20] After all, if we know the density of states then we can make an estimate of all the properties of the system employing the well-developed machinery of thermodynamics and statistical mechanics. There would arise no compulsive need for a Monte Carlo simulation. We may still decide to carry out Monte Carlo simulation, assemble an 'entropic' ensemble, and extract physical quantities employing un-weighting and re-weighting techniques.

[21] The set Ω shall contain all the microstates of the 'un-physical' system. These microstates can be of different energies. Let us group them in terms of their energies. Then we can say all the microstates of a group are equally probable and this probability is given by the

Nevertheless we shall consider Monte Carlo simulation of such an un-physical system employing the Metropolis algorithm and investigating the invariant probability density.

Let X_i be the current microstate in the Markov chain and $E_i = E(X_i)$ its energy. We have

$$p_i = p(X_i) \propto \frac{1}{\widehat{\Omega}(E_i)}.$$

Let X_t be the trial state and $E_t = E(X_t)$ its energy. We have

$$p_t = p(X_t) \propto \frac{1}{\widehat{\Omega}(E_t)}.$$

The probability of acceptance of the trial microstate is then given by

$$p = \min.\left(1, \frac{p_t}{p_i}\right) = \min.\left(1, \frac{\widehat{\Omega}(E_i)}{\widehat{\Omega}(E_t)}\right).$$

Note that if the trial microstate 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 towards low entropy region. This preference cancels statistically exactly the natural tendency of randomly sampling the trial microstates from the high entropy region. As a result, the Markov chain shall have an equal number of microstates in equal regions of energy. In other words, the energy histogram of the visited microstates 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 huge advantage because there are indeed energy barriers that emerge at temperatures close to the first-order phase transition and which are responsible for supercritical slowing of the dynamics. Also glassy systems have a free-energy profile with numerous ups and downs. Though we get an un-physical ensemble as a result of employing inverse of the density of states (in a Markov chain Monte Carlo method based on Metropolis rejection), the ensemble seems to possess certain desirable properties.

Of course we do not know yet the density of states. Perhaps it is a good idea to investigate further and invent methods that help us to obtain the density of states. If we embark on such an enterprise we may have

---

footnote continued

inverse of the density of states at that group energy. Each group would then constitute a microcanonical ensemble.

to forgo the comforts of Markov chain methodology and of the detailed balance present in the Metropolis rule. But then, we shall get easy access to entropy and other thermal properties, which have eluded the Markov chain Monte Carlo practitioners so far.

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 up to now, is tied crucially to our ability to assign a numerical value of the property O to every microstate of the system. Consider estimating a property like entropy. We cannot assign a numerical value for entropy to any single microstate of the system. All the microstates collectively own entropy. Hence thermal properties in general and entropy in particular are not easily accessible.

For computing thermal properties we need to go beyond Boltzmann Monte Carlo methods. That 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 multicanonical 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.

4. 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.

At the beginning of the simulation, define a function $g(E)$ and set it to unity for all E. Define also a histogram $H(E)$ and set it to zero for all E. Start with an arbitrary initial microstate X_0. Let $E_0 = E(X_0)$ be its energy. Update $g(E)$ and $H(E)$ as follows:

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

Here α is the Wang–Landau factor and we take $\alpha = e^1 = 2.7183$ in the first iteration. Generate a chain of microstates

$$X_0 \rightarrow X_1 \rightarrow \cdots \rightarrow X_i \rightarrow X_{i+1} \rightarrow \cdots$$

as per the algorithm described below.

Let X_i be the current microstate. Construct a trial microstate X_t. We need to decide whether to accept the trial state for advancing the chain. We take a decision on the basis of $g(E)$ updated at the end of the

previous step in which we selected the microstate X_i. Let $E_i = E(X_i)$ and $E_t = E(X_t)$. We have

$$p_i \propto \frac{1}{g(E_i)} \quad \text{and} \quad p_t \propto \frac{1}{g(E_t)}.$$

Define

$$p = \min.\left(1, \frac{p_t}{p_i}\right) = \min.\left(1, \frac{g(E_i)}{g(E_t)}\right).$$

The next microstate in the chain is

$$X_{i+1} = \begin{cases} X_t & \text{with probability} \quad p \\ \\ X_i & \text{with probability} \quad 1 - p \end{cases}$$

Once X_{i+1} is selected, the function $g(E)$ and the histogram $H(E)$ are updated. Carry out the simulation of the chain of microstates until the energy histogram becomes flat over, at least, a small range of energy. This constitutes one Wang–Landau iteration.

Note that the density-of-state-function $g(E)$ is updated at every step and the updated function is employed for decision making, from the very next step. As a result the chain of microstates generated, is not Markovian. The probability of transition between two microstates at any time step in the chain depends on how many times the chain has visited these two microstates in its past. The transition from the present to the future depends on the entire past. Hence we shall refer to the sequence of microstates as simply a chain and not prefix it with the adjective 'Markov'.

At the end of the first Wang–Landau iteration, change α to $\sqrt{\alpha}$. Reset $H(E)$ to zero for all E; but continue with $g(E)$. 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 α will move closer and closer to unity. For example, after some 25 iterations we shall have $\alpha = 1 + 3 \times 10^{-7}$. 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 $g(E)$ to the true but unknown density of states $\widehat{\Omega}(E)$. We take $g(E)$ obtained at the end of the last iteration – the one which generates a reasonably flat energy histogram – as an estimate of $\widehat{\Omega}(E)$, the true density of states.

We can define suitable criteria for measuring the flatness of the histogram. For example, we can consider the histogram to be flat if the smallest and the largest entries do not differ from each other by more than say

10%. Depending on 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 about how should it decreases (to unity) from one iteration to the next. The choice of $\alpha = \alpha_0 = e^1$, and the square-root rule, were recommended by Wang and Landau [12]. In principle, α_0 can be any real number greater than unity and it should decrease, preferably monotonically, to unity. Some researchers, see e.g. [13, 14], have found it advantageous to vary α non-monotonically at least initially. The important point is that any choice of variation of α that flattens the histogram would serve the purpose. In a sense, the histogram provides a diagnostic tool; you can monitor whether you are doing things right or wrong by looking at the flattening of the histogram. The flatter the histogram, closer is the density of states to its true value.

The Wang–Landau algorithm estimates the density of states only up to a normalization constant. In other words, the microcanonical entropy is estimated only up to 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. Your results are as good as the estimated density of states; but then, the density of states is accurate only if the energy histogram is flat. Any deviation from the flatness will lead to errors in density of states and hence errors in the estimated thermodynamic properties. We do not know yet how to translate the non-flatness in the histogram to errors in density of states and eventually to errors in the estimated thermodynamic properties.

5. Entropic ensemble

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

Let us call the set of microstates generated in the production run as an entropic ensemble or Wang–Landau ensemble. By employing un-weighting and re-weighting techniques we can make statically reliable

estimates of physical quantities[22]. The advantage is that the un-weighting followed by re-weighting to the desired ensemble will correct for errors in the density of states. Your final results will be right if the errors in density of states used for generating the entropic ensemble are reasonably small.

In what follows we shall show how to convert the entropic ensemble to a microcanonical ensemble and a canonical ensemble.

6. Entropic ensemble → microcanonical ensemble

Let

$$\{X_i : \ i = 1, 2, \ldots, M\}$$

denote a set of microstates belonging to the entropic ensemble. These microstates have been sampled from a probability distribution

$$p(X_i) \propto 1/g(E(X_i)).$$

Hence we first carry out un-weighting, see footnote (22):

$$W(X_i) = \frac{1}{1/g(E(X_i))} = g(E(X_i)).$$

Note that the microstates of the entropic ensemble are not necessarily of the same energy. In fact, the

---

[22]Let us explain un-weighting and re-weighting in a simple manner [5, 6]. Let x be a random variable and $f(x)$ its probability density. Let $h(x)$ be some function of x. The f-ensemble average of h is formally expressed as

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

Let $g(x)$ be a density function. Let us generate an ensemble $\Omega_g = \{x_i : \ i = 1, 2, \ldots, M\}$ by random sampling from $g(x)$. Our aim is to make an estimate of $\langle h \rangle_f$ employing the set Ω_g. Consider the following.

$$\langle h \rangle_f = \int_{-\infty}^{+\infty} dx\, h(x) f(x) = \int_{-\infty}^{+\infty} dx\, h(x) \frac{f(x)}{g(x)} g(x)$$

$$= \langle h\,(1/g)\, f \rangle_g.$$

The above is an exact result. The left-hand side is an f ensemble average of h. The right-hand side is a g ensemble average of h un-weighted by $1/g$ and re-weighted by f. The implementation goes as follows:

$$\langle h \rangle_f = \underset{M \to \infty}{\text{Limit}} \frac{1}{M} \sum_{i=1}^{M} h(x_i) \times \frac{1}{g(x_i)} \times f(x_i); \quad x_i \in \Omega_g.$$

ensemble contains an equal number of microstates in equal regions of energy – in other words the energy histogram is flat. For a microcanonical ensemble, all microstates are of the same energy and are equally probable. Hence the re-weighting factor is $1 \times \delta(E - E(X_i))$; the delta function ensures that we assemble only those microstates with the desired energy. Thus, we have

$$W(X_i) = g(E(X_i))\delta(E - E(X_i)).$$

Let $O(X_i)$ be the value of a property when the system is in the microstate X_i. The microcanonical ensemble average of O is given by

$$\langle O \rangle_{\mu C}(E) =$$

$$\underset{M \to \infty}{\text{Limit}} \frac{\sum_{i=1}^{M} O(X_i)g(E(X_i))\delta\,(E(X_i) - E)}{\sum_{i=1}^{M} g(E(X_i))\delta\,(E(X_i) - E)}.$$

In the above we have taken E as the energy of the isolated system described by the microcanonical ensemble.

Thus weighted averaging over microstates of a given energy belonging to the un-physical entropic ensemble equals averaging over a physical microcanonical ensemble modelling an isolated system.

7. Entropic ensemble → canonical ensemble

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

$$W(X_i) = g(E(X_i)) \times \exp[-\beta E(X_i)].$$

The canonical ensemble average of O is given by

$$\langle O \rangle_C = \underset{M \to \infty}{\text{Limit}} \frac{\sum_{i=1}^{M} O(X_i)g(E(X_i))\exp[-\beta E(X_i)]}{\sum_{i=1}^{M} g(E(X_i))\exp[-\beta E(X_i)]}.$$

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

From one single ensemble of microstates 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.

Thus, employing non-Boltzmann Monte Carlo methods, mechanical as well as thermal properties can be calculated. In what follows, we shall describe how does one calculate Landau free energy as a function of energy (or order parameter), for temperatures close to a transition point.

8. Calculation of free energy

The free energy of a closed system at temperature T is given by

$$F(T, V, N) = -k_B T \ln Q(T, V, N),$$

where $Q(T, V, N)$ is the canonical partition function. We can also define microcanonical free energy for an isolated system as

$$F(U, V, N) = U(S, V, N) - \left(\frac{\partial U}{\partial S}\right)_{V,N} S,$$

where $S(U, V, N)$ is the microcanonical entropy given by $S = k_B \ln \widehat{\Omega}(E, V, N)$.

For an equilibrium system, free energy cannot be a function of both energy and temperature simultaneously. An isolated system with a given energy has a unique temperature; a closed system at a given temperature has a unique (average) energy.

Let us say we are interested in estimating the penalty in terms of excess free energy required to keep a closed system in a state with its energy E different from the equilibrium energy $U(T)$. To this end, following Joon Chang Lee [15], we define a phenomenological free energy F_L, which is simultaneously a function of both energy and temperature. Let $F(T)$ denote the equilibrium free energy of the closed system at temperature T. We have $F_L(E, T) \ge F(T)$ for all energies, and equality obtains when $E = U(T)$. The excess free energy is given by

$$\Delta F = F_L(E, T) - F(T),$$

where

$$F_L(E, T) =$$

$$- k_B T \ln \sum_{C} g(E(C))\,\exp[-\beta E(C)]\,\delta\,(E(C) - E).$$

In the above, the sum runs over all the microstates belonging to the entropic ensemble collected during the production run of the Wang–Landau Monte Carlo simulation.

8.1 *Phase transition in a q-state Potts spin model with q = 8: Free energy versus energy*

To illustrate the method for calculating free energy as a function of temperature and energy/order parameter, we present the results on q-state Potts model [16, 17]. In this model, a q-state spin variable,

$$\sigma_i = 0, \, 1, \, 2, \, \ldots, \, q-1$$

is placed at each lattice site. The Hamiltonian for the spin–spin interaction is given by

$$H = -J \sum_{\langle i,j \rangle} \delta_{\sigma_i,\sigma_j},$$

where δ is the Kronecker delta function:

$$\delta_{x,y} = \begin{cases} 1 & \text{if } x = y \\ 0 & \text{if } x \neq y \end{cases}$$

and $\langle i, j \rangle$ denotes the nearest-neighbour sites on the lattice. The energy of two neighbouring spins is $-J$ if they are in the same state and zero otherwise. J is a positive constant and measures the strength of ferromagnetic interaction between two neighbouring spins. Without the loss of generality we have taken $J = +1$. The ground state is q-fold degenerate.

In two dimensions, the Potts model exhibits a continuous phase transition (second order) for $q \leq 4$, whereas the transition is discontinuous (first order) for $q > 4$.

We have considered a two-dimensional square lattice of size 20×20; each lattice site holds a Potts spin with $q = 8$. The system exhibits a first-order phase transition. Initially we start with a random configuration and employing the Wang–Landau algorithm we estimate the density of states. We then calculate the specific heat (C_v) as a function of temperature; the value of temperature at which C_v is maximum is taken as the transition point, T_c. We obtain Landau free energy profiles at T_c and temperatures below and above T_c as explained in the previous section. We find that at $T = T_c = 0.749$ the specific heat is maximum.

Figure 1 shows $F_L(E, T)$ vs. E for $T = 0.748, 0.749$ and 0.750. It is clear from these figures that the transition is of the first order.

8.2 *Melting of hairpin-DNA (hpDNA)*

The next problem we have chosen for the purpose of illustration is melting transition of hpDNA. It is known that a DNA has two strands of equal length twisted around a common axis to form a double helix. Each strand can be considered as a chain of monomers or bases. When complementary bases at the two ends of a strand pair up, a hairpin structure occurs, with a loop and a stem. The structure can be closed or open. The transition from the closed to open state is called denaturation or unbinding. If this denaturation occurs by virtue of temperature, then it is called melting.

Here, we present results on melting of hpDNA with a chain length of 100 units. We estimate the density of states by employing the Wang–Landau algorithm. We have considered four-site occupation, bond fluctuation model [18], on a two-dimensional square lattice in our simulation. For details of the simulation, see [19]. Next we calculate heat capacity as a function of temperature and determine the temperature at which it is maximum. We take this as the transition temperature T_c. Then we obtain the free energy vs. energy profiles and these are depicted in figure 2, for $T < T_c$, $T = T_c$ and $T > T_c$, respectively. These free energy profiles clearly show that the transition is of first order.

9. Free energy vs. order parameter

The Landau free energy is usually expressed as a function of order parameter for temperatures infinitesimally close to transition point on either side. Let M denote the order parameter. To include the order parameter in the calculations, Landau, Tsai and Exler [20] defined a joint density of states $g(E, M)$ and employed it in the Wang–Landau algorithm. Let E_i and M_i be the energy and order parameter of the current microstate; let E_t and M_t denote the same quantities for the trial microstate. The acceptance probability in the Metropolis rejection step is taken as

$$p = \min \left(1, \frac{g(E_i, M_i)}{g(E_t, M_t)} \right).$$

The joint density of states is updated at every step. We also monitor the histogram $H(E, M)$. The histogram now is a two-dimensional surface. When the surface becomes flat we can say that $g(E, M)$ has converged to the true two-dimensional density of states, $\widehat{\Omega}(E, M)$. Obtaining a flat histogram in two-dimensional space requires a considerable computing time [20].

An alternate and better strategy proposed in [21] involves estimating the two-dimensional density states in the production run (see below). The learning run is reserved, as usual, for obtaining a converged one-dimensional density of states, $g(E)$, in energy space only.

We assemble, in the production run, an entropic ensemble. We carry out un-weighting of the microstate and estimate the joint density of states as an average
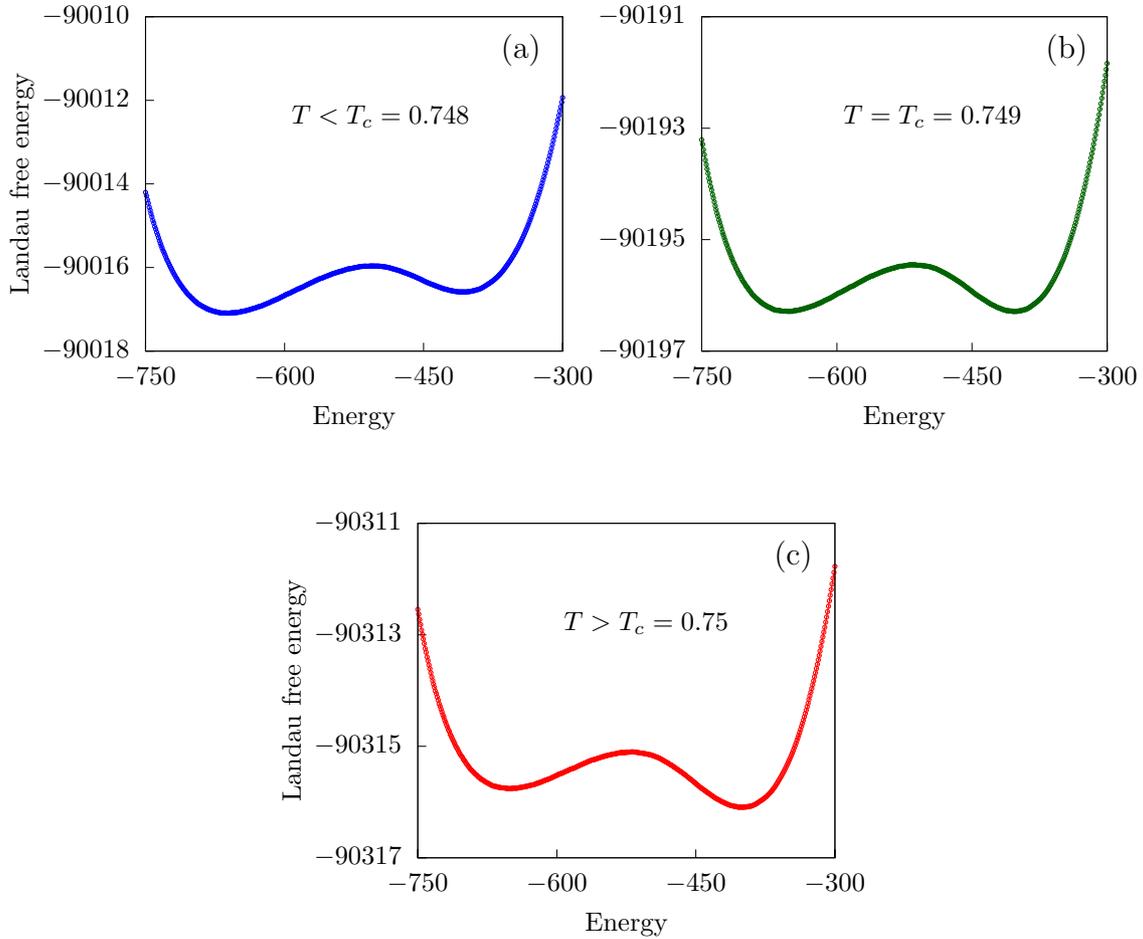
Figure 1. Landau free energy as a function of energy for $q = 8$ states Potts model on a two-dimensional square lattice of size $L = 20$. (**a**) is for $T = 0.748 < T_c$; (**b**) is for $T = 0.749 = T_c$ and (**c**) is for $T = 0.750 > T_c$.

over the resulting microcanonical ensemble.

$$\langle g(E_i, M_j) \rangle =$$

$$\frac{\sum_C \delta(E(C) - E_i)\, \delta(M(C) - M_j)\, g(E(C))}{\sum_C \delta(E(C) - E_i)\, g(E(C))}$$

$$\forall \quad E_i, M_j.$$

In the above, the sum runs over all the microstates of the entropic ensemble.

Having estimated the joint density of states, we calculate the Landau free energy as

$$F_L(M_j, T) = -k_B T \ln \sum_i \langle g(E_i, M_j) \rangle \exp(-\beta E_i),$$

where we take T to be very close to the transition temperature.

9.1 $q = 2$ *state Potts model*

As mentioned earlier, the q-state Potts model on a two-dimensional lattice exhibits a continuous phase transition for $q < 4$. Hence we consider $q = 2$ and a 20×20 square lattice[23]. Figure 3 depicts the Landau free energy profiles for Potts spin model with $q = 2$.

10. End note

We started this review with a discussion on the Metropolis algorithm to sample microstates 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

---

[23] Note that Potts spin model with $q = 2$ is similar to the well-known Ising model. In the Ising model, we have spins with $S = \pm 1$ whereas in the $q = 2$ Potts model, we have spins with $\sigma = 0, 1$.
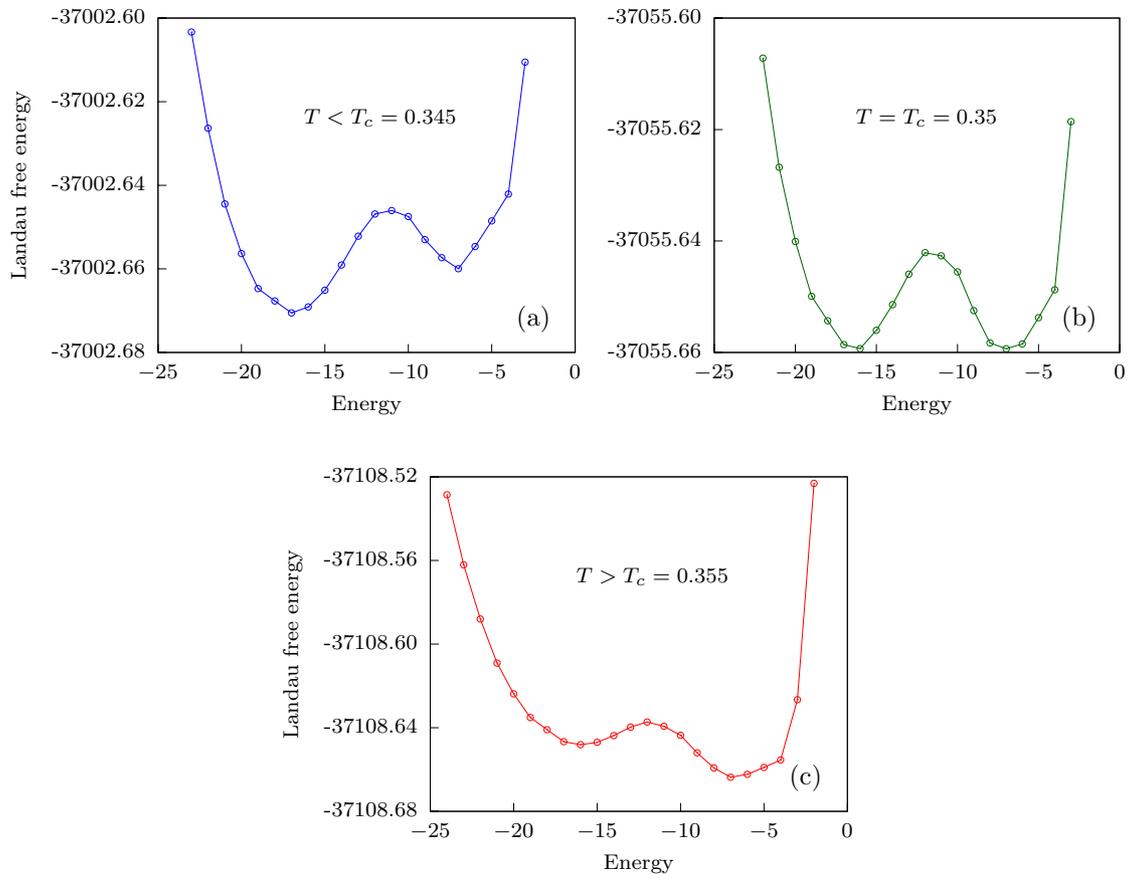
Figure 2. Variation of free energy with energy (**a**) at $T < T_c$, (**b**) at $T = T_c$ and (**c**) at $T > T_c$ for a chain of length of 100.
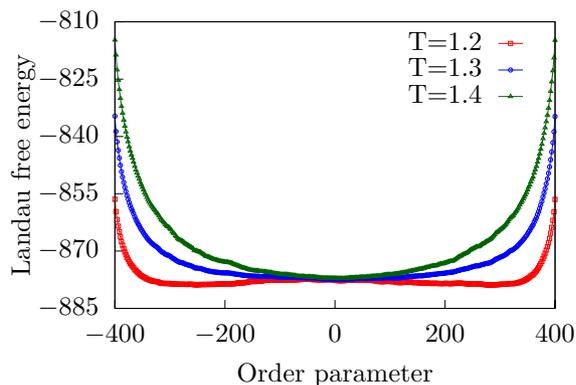

Figure 3. Landau free energy as a function of order parameter (magnetization) for $q = 2$ states Potts model.

energy. The reason is simple. A value for a mechanical property can be assigned to each microstate.

However, Boltzmann sampling is quite clumsy when it comes to estimating thermal properties like entropy and free energies. The clumsiness owes its origin to the fact that a numerical value for entropy cannot be assigned to any single microstate. All the microstates, collectively, own entropy. Entropy is a property of an ensemble and not of any single microstate. This problem about estimating entropy was recognized even in the early days of Monte Carlo practice by Torrie and Valleau [9]; they invented umbrella sampling which addresses these issues. Umbrella sampling has since inspired and given rise to a whole lot of non-Boltzmann methods; the latest to arrive is the method of Wang and Landau [12]. We have reviewed the basic idea behind the Wang–Landau algorithm and described 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 better, for calculating the mechanical properties. Besides, they provide reliable estimates of thermal properties, not easily accessible to Boltzmann Monte Carlo methods.

We must quickly add that all is not cosy about the Wang–Landau algorithm. There are issues and there are difficulties. A typical Monte Carlo aficionado, see e.g. [22], does not feel comfortable since the algorithm does not obey the detailed balance; in fact, the chain

generated in the learning run is not even Markovian. What, then, guarantees convergence of $g(E)$ to $\widehat{\Omega}(E)$?

Also the algorithm performs poorly on systems with continuous degrees of freedom. There is a slowing down of dynamics but now due to entropy barriers. These and related problems have attracted the attention of several researchers, see e.g. [13, 14, 23–25] and remedies have been suggested. But in my opinion no satisfactory solution has yet emerged. All the remedies suggested so far seem ad-hoc.

There are also issues about error – 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, the pride of a Monte Carlo practitioner lies often in his ability to compute not only averages, but also associated statistical errors. But then we do not know how to calculate Monte Carlo error bars in a Wang–Landau simulation.

We hope these and other issues would get resolved satisfactorily soon and non-Boltzmann Monte Carlo methods would emerge as standard tools in the arsenal of a Monte Carlo practitioner. With this optimistic note, let us close this review.

Acknowledgements

References

[1] R K Pathria, *Statistical Mechanics*, 2nd edn (Butterworth, Heinemann, 1996)

[2] N Metropolis, A W Rosenbluth, M N Rosenbluth, A H Teller and E Teller, *J. Chem. Phys.* **21** 1087 (1953); see also G Bhanot, *Rep. Prog. Phys.* **51**, 429 (1988)

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

[4] A Papoulis, *Probability 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 (2000); see arXiv:cond-mat/014215v1, 12 April 2001

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

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

[9] G M Torrie and J P Valleau, *J. Comput. Phys.* **23**, 187 (1977)

[10] B A Berg and T Neuhaus, *Phys. Rev. Lett.* **68**, 9 (1999)

[11] J Lee, *Phys. Rev. Lett.* **71**, 211 (1993); *Erratum*, **71**, 2353 1993

[12] F Wang and D P Landau, *Phys. Rev. Lett.* **86**, 2050 (2001)

[13] D Jayasri, V S S Sastry and K P N Murthy, *Phys. Rev. E* **72**, 36702 (2005)

[14] P Poulin, F Calvo, R Antoine, M Broyer and P Dugord, *Phys. Rev. E* **73**, 56704 (2006)

[15] J C Lee *Thermal Physics: Entropy and Free Energies* (World Scientific, Singapore, 2002)

[16] R Potts, *Math. Proc. Cambridge Philos. Soc.*, **48(1)**, 106 (1952)

[17] F Y Wu, *Rev. Mod. Phys.* **54** 235, (1982)

[18] I Carmesin and K Kremer, *Macromolecule* **21**, 2819 (1988)

[19] M Suman Kalyan and K P N Murthy, *Physica A* **428**, 38 (2015)

[20] D P Landau, S-H Tsai and M Exler, *Am. J. Phys.* **72**, 1294 (2004)

[21] M Suman Kalyan, R Bharath, V S S Sastry and K P N Murthy, *J. Stat. Phys.* **163**, 197 (2016)

[22] 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, volume 3, ed. Y Holovatch (World Scientific, 2012), pp. 93–166

[23] B J Schulz, K Binder, M Müller and D P Landau, *Phys. Rev. E* **67**, 67102 (2003)

[24] C Zou and R N Bhatt, *Phys. Rev. E* **72**, 25701(R) (2005)

[25] A Tršter and C Dellago, *Phys. Rev. E* **71**, 66705 (2005)