1

Bias Monte Carlo Methods in Environmental Engineering

Albert S. Kim
Civil and Environmental Engineering, University of Hawaii at Manoa
USA

1. Introduction

In the 21st century, vital resources for human beings such as food, energy, and water (FEW) are being rapidly depleted. Global water scarcity has already become a serious world-wide problem; "the cheap energy" – fossil fuels – will last only a few decades; and skewed global food distributions are marked by serious obesity in one region and deadly starvation in another. The standard role of environmental engineering now vigorously extends from providing conventional sanitation guidelines to contributing crucial information to environmental policy-making and futurological issues. Unlike other engineering and closely related disciplines (such as chemical engineering, electrical engineering, material science, and computer and information sciences), environmental engineering deals with poorly or incompletely defined problems whose scientific origins are in multiple sub-categories of physics, chemistry, biology and mathematics; and spontaneously gives birth to novelties in interdisciplinary research areas.

In general, medicine is classified into curative and preventive technologies. One can make an analogy of the curative medicine (alleviating pain from diseases) to conventional engineering that tries to improve the quality of human life. Preventive medicine is analogous to the corresponding role of environmental engineering which, in part, is to conserve the natural environment by eliminating or minimizing environmental risks. The term "conservation" is often regarded as passive human responses to return a degraded system to its original state after accidents. Prevention refers to keeping something from unexpected happening or arising. Active engineering responses can include designing new paradigms for environmentally friendly, green, or zero-emission processes to eliminate potential adverse effects on nature from undesired technological by-products.

A question arises to researchers in environmental engineering and science,

> "*What do we prevent and how, if we do not know what is really happening?*"

Perhaps this question may be a senseless one, if one develops new products such as cellular phones, computer chips, or sign-recognition software, because market demands truly control developers' objectives to generate better money-making commercial products. As an environmental engineer, how do we choose *probably* the most urgent and long-term-impact problems; and then clearly define *probably* beneficial outcomes for human beings by solving the uncertain problems? Considering these questions and above issues, don't we have enough reasons to deal with fundamentals of probability and statistics and see how these are used in thermodynamics in order to deeply understand natural and engineered phenomena? *What are the likelihood, chances, and probabilities in nature?*

1.1 Probability

The primary objective of this chapter is to introduce how to use statistical mechanics to deal with engineering problems, specifically with environmental engineering applications. In physics, subjects of conservation laws include mass-energy, (translational and rotational) momentum, electric charge, and (more importantly?) probability, which are kept constant during the time in a closed system. Excluding conversion between mass and energy which rarely happens in practical engineering processes, one can explain energy conservation, indicating total energy as the sum of the kinetic energy and potential energy, which is always the same number. However, one should notice that this relationship stems from Newton's second law and energy is nothing but a constant generated by integrating

$$\boldsymbol{F} = m\boldsymbol{a} = -\nabla V(\boldsymbol{r}) \tag{1}$$

with respect to the object's position $\boldsymbol{r}$ in a conservative field where $\boldsymbol{F}$ is a force acting on an object, m is the object mass, $\boldsymbol{a}$ is the acceleration, and $V(\boldsymbol{r})$ is the potential energy. In one dimension, multiplying v on both sides of Eq. (1) yields

$$mv\frac{\mathrm{d}v}{\mathrm{d}t} = -\frac{\mathrm{d}x}{\mathrm{d}t}\frac{\mathrm{d}V}{\mathrm{d}x} \tag{2}$$

which is integrated as

$$m\int_{v_1}^{v_2} v\mathrm{d}v = -\int_{x_1}^{x_2} \mathrm{d}V \tag{3}$$

$$\frac{1}{2}mv_2^2 - \frac{1}{2}mv_1^2 = -V(x_2) + V(x_1) \tag{4}$$

assuming that the particle of mass m is at x_1 having velocity v_1 at time $t = t_1$ and moves to x_2 having velocity v_2 at time $t = t_2$. Thus,

$$\frac{1}{2}mv_1^2 + V(x_1) = \frac{1}{2}mv_2^2 + V(x_2) = E = \text{Constant} \tag{5}$$

which implies that the sum of kinetic and potential energies is a constant, denoted as E. Feynman et al. (1963) indicated that

> "It is important to realize that in physics today, we have no knowledge of what energy *is*. We do not have a picture that energy comes in little blobs of a definite amount. It is not that way. However, there are formulas for calculating some numerical quantity, and when we add it all together it gives '28' – always the same number. It is an abstract thing in that it does not tell use the mechanism or the *reasons* for the various formulas."

As noted above, we do not know what energy is exactly but we know that it has a constant character with respect to time. Similar to Feynman's description about energy, probability is assumed to be implicitly understood by readers; and sometimes a more non-technical word, "chance", is alternatively and widely used. If a sales person signed an important contract today, he might ask to himself: What is the chance that I will get a promotion call from my boss next week? A Hawaii politician might be interested in the question: What is my chance of being elected mayor of Honolulu? These chances are hard to estimate because the questioners do not have enough information, and the knowledge is sometimes too uncertain to be quantified.

Without exception, the probability is a fraction, i.e., a dimensionless number between 0 and 1, often measured as percentage. The widely used examples in the study of probability include tossing a coin and rolling a dice, and the following questions are often asked: What is the

Number	1	2	3	4	5	6
Regular	1/6	1/6	1/6	1/6	1/6	1/6
Modified	0/6	1/6	1/6	1/6	2/6	1/6

Table 1. Probability distribution of the regular and modified dice.

probability of heads when tossing a coin?; and, what is the probability getting a 2 when rolling a dice?. The answers to the first and second questions are 1/2 and 1/6, respectively, which almost nobody refutes. But, why? More specifically, why do heads and tails have the same chances of 1/2? Or, why do the six consecutive numbers of a dice have the same probability of 1/6 to be thrown? A simple answer is that we assigned an equal chance to all probable outcomes. If so, again, why do we do that? Perhaps, it is because we do not have any better knowledge than that of the equal distribution. This was called the "principle of incomplete reasons" (PIR). A similar concept, "equipartition principle", can be found in statistical mechanics, i.e, putting exactly 1/3 of the total energy to each direction of homogeneous and isotropic three-dimensional space. Note that homogeneity and isotropy are also our great assumptions.

These are excellent pedagogical examples but truly ideal. What if someone has a damaged coin that is not flat enough, so that we cannot convince ourselves of the equal probabilities of heads and tails? What if someone added four more dots on the surface of the dice with one dot and now it displays 5 (see Figure 1), but we keep throwing the dice without knowing about the significant change in the probability distribution. Then, sample space, technically called "ensemble", was modified, so that the uniform distribution fails to statistically describe the system.

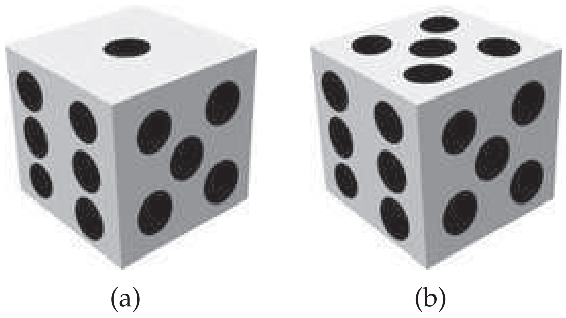

(a) (b)

Fig. 1. Dice with (a) 1 to 6 and (b) 1 replaced by 5

Now, heads and tails have different probabilities of occurrence due to the uneven shape; and by tossing the modified dice, the probability of throwing 1 is zero, and that of 5 is 2/6, not 1/6. How do we evaluate the expected values of the two examples after the changes? The dice case would be easier to reconstruct the probability distribution as shown in Table 1. After a large number of tosses, the average outcome, i.e., expectation value, is

$$1 \times \frac{0}{6} + 2 \times \frac{1}{6} + 3 \times \frac{1}{6} + 4 \times \frac{1}{6} + 5 \times \frac{2}{6} + 6 \times \frac{1}{6} = \frac{25}{6} = 4.167$$

Note that we still use the principle of incomplete reasons by assigning equal probability of 1/6 to rolling 2, 3, 4 or 6; and moving the probability 1/6 from 1 to 5. The expectation value of the regular dice is 3.5, which is similar to 4.167. If our number of tosses is not large enough, then the modified probability distribution would not be achieved and the expectation value can be accepted within a reasonable(?) range of tolerance error. In the above case of the modified dice, we implicitly assume that the shape is a regular cubic. When a coin is damaged, i.e., curved and/or stretched, then the uneven probability distribution must be made even and is not easy to build.

We do not have enough information at the micro-mechanics level from which we can definitely say that "The probability of tossing heads on the damaged coin is $\sqrt{2}/2$ and that of tails is $1-\sqrt{2}/2$." There are many important features in tossing the damaged coin: tosser's specific way of flipping the coin into air, the number of spins before landing, the landing conditions such as falling velocity and bouncing angle, all of which were ignored for the undamaged coin. To average out these specific impacts on the probabilities of throwing heads or tails using the damaged coin, we need to have the number of tossings much more than that of tossing a regular coin. In other words, we cannot equally distribute the probabilities to heads or tails, and we do not know how much the two probabilities are different from 1/2. So, we do a large number of tossing experiments to estimate probabilities of landing on heads and tails, keeping the fact that a sum of the two equal probabilities is always 1, no matter how much the coin was damaged. Note that we used our basic belief of equal probability distribution for tossing the modified dice; and on the other hand, we actually did a number of tossing experiments of the damaged coin to estimate the probability distribution. This is because it is mathematically formidable to calculate probability distribution of the damaged coin; and we still believe the probabilities to throw 2, 3, 4 and 6 of the modified dice are equally 1/6.

1.1.1 Conventional point of view: Frequency

Reif (1965) indicated in his book that the probability of the occurrence of a particular event is defined with respect to a particular ensemble consisting of a very large number of similarly prepared systems; and is given by the fraction of systems in the ensemble which are characterized by the occurrence of the specific event. Therefore, the fraction, called *frequency probability*, is the ratio of a certain occurrence of our interest to the total number of possible occurrences. This surely implies that unless we do a large number of experiments, the measured frequency fraction is not accurate enough. In general, spanning all the possible cases is a formidable task, especially for a complex system.

1.1.2 Bayesian Point of view: Distribution

An alternative approach is the *conditional probability*, which allows one to localize the sampling space and provide a new probability distribution. The probability that both a and b occur is expressed as

$$P(a \cap b) = P(a|b)\,P(b) = P(b|a)\,P(a) \tag{6}$$

indicating that $P(a \cap b)$ is equal to

1. the probability of a occurring, $P(a)$, times the probability of b occurring given a has occurred, $P(b|a)$, and
2. the probability of b occurring, $P(b)$, times the probability of a occurring given b has occurred, $P(a|b)$.

so that $P(a|b)$ is written as

$$P(a|b) = \frac{P(b|a)\,P(a)}{P(b)} \tag{7}$$

The proof for Eq. (6) uses two basic probability relationships: the sum rule and the product rule, i.e.,

$$P(a|b) + P(\widetilde{a}|b) = 1 \tag{8}$$

and

$$P(a \cap a'|b) = P(a|b)P(a'|b) \tag{9}$$

respectively. Eq. (8) indicates that, given b has occurred, the sum of probabilities of a occurring, $P(a|b)$, and not a occurring (i.e., $\widetilde{a}$), $P(\widetilde{a}|b)$, is equal to 1. The product rule of Eq. (9) means the probability of occurring a and a' given that b has occurred, $P(a \cap a'|b)$, is equal to the probability of a, given b, $P(a|b)$, multiplied by the probability of a' given b, $P(a'|b)$.

1.1.3 Examples

In this section, well-known examples are selected and solved using the frequency and conditional probabilities. In addition, logical ways of solving the example problems are included. The purpose of this section is to show that the conditional probability method is as powerful as the other two methods.

A. Monty Hall dilemma

Suppose you're on a game show, and you're given the choice of three doors. Behind one door is a car; behind the others, goats. You pick a door, say #1, and the host, who knows what's behind the doors, opens another door, say #3, which has a goat as behind it. He then says to you,

> "Do you want to switch to door #2, or stay with door #1?"

Is it to your advantage to change your choice? It is better to stay with door #1 or is it better to switch to door #2 or is the probability of winning the same for either choice?

Fig. 2. The Monty Hall paradox first appeared in 1975 on the American television game show *Let's Make a Deal*, hosted by Mr. Monty Hall (1921 – present). The game show aired on NBC daytime from December 30, 1963, to December 27, 1968, followed by ABC daytime from December 30, 1968, to July 9, 1976, along with two primetime runs. It also aired in syndication from 1971 to 1977, from 1980 to 1981, from 1984 to 1986, and again on NBC briefly from 1990 to 1991. Historical records from Wikipedia (`http://en.wikipedia.org/wiki/Monty_Hall`) and special thanks to Tae Chun for the illustration.

(a) Solution using logical thinking

When you selected door #1, the probability of winning the car is 1/3. No question at all! But there are also two other doors. The host opened door #3, showing a weird-looking goat. This makes the original probability assigned to door #3 equal to zero. Where has it gone since the sum of the probability of all possible events should always be one. You don't think the winning probability on door #1 has changed. Then, there is only one possibility, i.e., the probability of door #3 moved to that of door #2. So, if you switch to door #2, your winning chance will be doubled: from 1/3 to 2/3. So, you are switching now!

(b) Solution using conditional probability

Perhaps the logical solution above might not be clear enough. So let's calculate the conditional probability using Bayes' rule (Bayes and Price, 1763). The game show can be mathematically described as three sets with possible cases:

- S = my Selection = {1, 2 or 3}
- H = Host open = {1, 2 or 3} $\notin S$
- C = Door for car = {1, 2 or 3} $\notin H$

You want to know the probability of winning after switching from door #1 to #2. Without loosing generality, this probability can be written as $P\left(C_2|S_1 \cap H_3\right)$, which is the probability that the car is behind door #2 given that you selected door #1 and the host opened door #3 showing a goat (not a car!). By substituting $a = C_2$ and $b = S_1 \cap H_3$ into Eq. (7), one can write in a symmetric form

$$P\left(C_2|S_1 \cap H_3\right) = \frac{P\left(S_1 \cap H_3|C_2\right) P\left(C_2\right)}{P\left(S_1 \cap H_3\right)} \tag{10}$$

of which each probability can be addressed as follows.

First, $P\left(C_2\right)$ is the probability that the car is behind door #2, which is equal to

$$P\left(C_2\right) = \frac{1}{3} = P\left(C_1\right) = P\left(C_3\right) \tag{11}$$

because the probability of finding the car is equally distributed among the three doors. This resembles the energy equipartition principle.

Second, $P\left(S_1 \cap H_3|C_2\right)$ is the probability of S_1 and H_3 given C_2 so substituting into Eq. (9) yields

$$P\left(S_1 \cap H_3|C_2\right) = P\left(S_1|C_2\right) P\left(H_3|C_2\right) \tag{12}$$

where C_2 confines a sub-domain of probability for S_1 and H_3. Because we do not know which door will reveal the car, our first selection of a door is independent of the probability of the car being behind door #2:

$$P\left(S_1|C_2\right) = P\left(S_1\right) = \frac{1}{3} \tag{13}$$

However, the host knows that the car is behind door #2 and he also saw that you selected door #1. Therefore, given C_2 (that the host is aware of), the probability that the host opens door #3 is

$$P\left(H_3|C_2\right) = 1 \tag{14}$$

so that

$$P\left(S_1 \cap H_3|C_2\right) = \frac{1}{3} \times 1 = \frac{1}{3} \tag{15}$$

Third, you need to calculate $P\left(S_1 \cap H_3\right)$, the probability that S_1 and H_3 (and vice versa) will happen, which is simply equal to the probability of S_1 multiplied by the probability of H_3, i.e.,

$$P\left(S_1 \cap H_3\right) = P\left(S_1\right) P\left(H_3\right) = \frac{1}{3} \cdot \frac{1}{2} = \frac{1}{6} \tag{16}$$

because we select one door out of three and the host opens one out of the two remaining doors.

Finally, the winning probability after switching from door #1 to #2 is calculated as

$$P\left(C_2 | S_1 \cap H_3\right) = \frac{P\left(S_1 \cap H_3 | C_2\right) P\left(C_2\right)}{P\left(S_1 \cap H_3\right)} = \frac{\dfrac{1}{3} \cdot \dfrac{1}{3}}{\dfrac{1}{6}} = \frac{2}{3} \tag{17}$$

and the winning probability by staying with door #1 is calculated using the the sum rule:

$$P\left(C_1 | S_1 \cap H_3\right) = 1 - P\left(C_3 | S_1 \cap H_3\right) - P\left(C_2 | S_1 \cap H_3\right) = 1 - 0 - \frac{2}{3} = \frac{1}{3} \tag{18}$$

Note that the probability of the car being behind door #3 after the host opened #3 is zero, i.e., $P\left(C_3 | S_1 \cap H_3\right) = 0$. This indicates if you stay in door #1, then the probability of winning the car is 1/3, i.e., $P\left(C_1 | S_1 \cap H_3\right) = 1/3$, but if you switched to door #2, the probability is doubled! *So, always switch your door!* Additional analysis of the conditional probability can be found in section 5.1.

(c) Solution using frequency probability
Let's assume that the car is behind door #1.

1. If you select door #1, the host will open either door #2 or #3. Let's say, door #3. If you switch your door, then you won't get the car.
2. If you select door #2, there is no question at all that the host will open door #3. If you switch from door #2 to #1, then you will win the car.
3. If you select door #3, the host will open door #2. If you switch from door #3 to #1, then you will win the car.

So among the three possible cases above with unconditional switch no matter which door is selected, two cases give car-winning opportunities. Therefore, the probability of winning the car by switching to the other door is 2/3. This solution method seems to be easier than that of the conditional probability above, but building a complete sample space is not always easy.

B. Prisoner's Dilemma

This example is taken from a book written by Mosteller (1965). "Three prisoners, A, B, and C, with apparently equally good records have applied for parole. The parole board has decided to release two of the three, and the prisoners know this but not which two. A warder friend of prisoner A knows who will be released. Prisoner A realizes that it would be unethical to ask the warder if he, A, is to be released, but thinks of asking for the name of *one* prisoner *other than himself* who is to be released. He thinks that before he asks, his chances of release are $\frac{2}{3}$. He thinks that if the warder says "B will be released," his own chances have now gone down to $\frac{1}{2}$, because either A and B or B and C are to be released. And so A decided not to reduce his chances by asking. However, A is mistaken in this calculations. Explain."

(a) Solution using logical thinking
The probability that A will be released is $\frac{2}{3}$ because two out of the three will be released. The decision of the parole board is independent of A's knowledge. Therefore, A still has a 2/3 chance of being released.

(b) Solution using conditional probability

The probability of A being released, given that B will be released, can be expressed as

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{19}$$

One calculates $P(A) = 2/3$, $P(B) = 1$, and $P(B|A) = P(B) = 1$ because A does not affect B. Therefore,

$$P(A|B) = \frac{1 \cdot \frac{2}{3}}{1} = \frac{2}{3} \tag{20}$$

indicating that no matter whether A knows about B's fate or not, the probability of A's release is $2/3$.

(c) Solution using frequency probability

The possible pairs to be released are AB, BC, and AC, which have equal probability of $2/3$. Then, the probabilities of possible cases in the sample space are calculated as

Released	Warder says	Probability
AB	B	$1/3$
AC	C	$1/3$
BC	B	$1/6$
BC	C	$1/6$

Thus, the probability of A being released is equal to

$$\frac{\text{Probability of } AB \text{ to be released}}{\text{Probability of } AB \text{ to be released + Probability of } BC \text{ to be released}}$$

given that B will be released. Therefore, A's probability of being released is

$$\frac{\frac{1}{3}}{\frac{1}{3} + \frac{1}{6}} = \frac{2}{3} \tag{21}$$

As shown above in the two examples, conditional probability is as powerful as frequency probability and has mathematical elegance. Now we will see how conditional probability is efficiently used in statistical physics when dealing with a large population.

1.2 Thermodynamics and statistical mechanics

1.2.1 Heat and work

Statistical mechanics, as a branch of theoretical physics, studies macroscopic systems from a microscopic or molecular point of view, dealing with systems in equilibrium. It is often referred to as *statistical thermodynamics* as it links (classical) thermodynamics with molecular physics. Thermodynamic laws describe the transport of heat and work in thermodynamic processes.

- The 0^{th} law of thermodynamics: If two thermodynamic systems are each in thermal equilibrium with a third, then they are in thermal equilibrium with each other. In other words, if $A = B$ and $B = C$, then $A = C$.
- The 1^{st} law of thermodynamics: Energy is neither created nor destroyed. Increase in the internal energy E of a system is equal to the heat Q supplied to the system subtracted by the work W done by the system, i.e., $\text{d}E = \text{đ}Q - \text{đ}W$. The symbol "đ" indicates that the

differential is inexact. Q and W are path functions, and E is a state function. Specifically in equilibrium, $\mathrm{d}E = T\mathrm{d}S - P\mathrm{d}V$, where the temperature T and pressure P are integral factors of $\mathrm{d}S$ and $\mathrm{d}P$, respectively.

- The 2^{nd} law of thermodynamics: Spontaneous natural processes increase entropy overall. In other words, heat can spontaneously flow from a higher-temperature region to a lower-temperature region, but not the other way around:

$$\Delta S = S_T - S_0 = \int_{T_0}^{T} \frac{dQ}{T} \geq 0 \tag{22}$$

  where $\Delta S = 0$ is for the reversible process.

- The 3^{rd} law of the thermodynamics: As the temperature approaches absolute zero, the entropy of a system approaches a constant minimum. Briefly, this postulates that entropy is temperature dependent and results in the formulation of the idea of absolute zero. At $S_0 = 0$, T is defined at 0K.

The heat absorbed by the system from the surroundings during the change from state A to state B is

$$Q = \int_A^B \text{đ}Q = \int_A^B T\mathrm{d}S \tag{23}$$

where T is the absolute temperature and S is the entropy. The pressure-volume work done by a thermodynamic system on its surroundings that goes from state A and state B is

$$W = \int_A^B \text{đ}W = \int_A^B P\mathrm{d}V \tag{24}$$

where P is the pressure exerted by the surroundings on the system and $\mathrm{d}V$ is an infinitesimal change in volume. The work Q and heat W have different values for different paths from state A to B so that Q and W are *path functions*. However, the first law of thermodynamics states that the infinitesimal difference between Q and W is independent of the path, i.e.,

$$\mathrm{d}E = \text{đ}Q - \text{đ}W \tag{25}$$

where E is a state function and called the internal energy. The second integrals in Eqs. (23) and (24) are valid for *reversible* processes in which there exist integral factors: T for $\text{đ}Q$ and P for $\text{đ}W$. Therefore, $\mathrm{d}S$ and $\mathrm{d}V$ are exact differentials: S and V are *state* functions. Thus,

$$\Delta S = \int_A^B \frac{\text{đ}Q}{T} \geq 0 \tag{26}$$

where the equals sign is for a reversible process. Eq. (26) indicates the second law of thermodynamics. In an irreversible process, the entropy of the system and its surroundings increase; in a reversible process, the entropy of the system and its surroundings remains constant. In other words, the entropy of the system and its surroundings never decreases! The third law of thermodynamics allows us to calculate the absolute entropy of a substance:

$$S - S_0 = \int_0^T \frac{\text{đ}Q}{T} \tag{27}$$

where $S_0 = 0$ at $T = 0$K. For simple systems, the first law of Eq. (25) can be expressed as

$$\mathrm{d}E = T\mathrm{d}S - P\mathrm{d}V \tag{28}$$

1.2.2 Microstates in phase space

The number of possible cases that N particles exist in m distinct microstates in phase space is

$$W = \frac{N!}{n_1!\,n_2!\,\cdots\,n_i!\,\cdots\,n_m!} = \frac{N!}{\prod_{i=1}^{m} n_i!} \tag{29}$$

where n_i is the number of particles in state i running from 1 to m so that

$$\sum_{i=1}^{m} n_i = N \tag{30}$$

or

$$\sum_{i=1}^{m} f_i = 1 \tag{31}$$

where frequency f_i is defined as

$$f_i = \frac{n_i}{N} \tag{32}$$

Usually, N and n_i are large numbers, which allow us to use Stirling's formula (Reif, 1965):

$$x! \approx x \ln x - x \tag{33}$$

to obtain

$$\begin{aligned} \frac{1}{N}\ln W &= \frac{1}{N}\ln N! - \frac{1}{N}\ln\left[\prod_i n_i!\right] \\ &\approx \ln N - 1 - \frac{1}{N}\sum_i (n_i \ln n_i - n_i) \\ &= -\sum_i^m f_i \ln f_i \end{aligned} \tag{34}$$

The system energy can be expressed as the sum of the product of the energy of state i and the number of particles in the state:

$$\sum_{i=1}^{m} f_i \epsilon_i = E \tag{35}$$

Our goal is to find function f_i that maximizes $(\ln W)/N$ with the two constraints of Eqs. (30) and (35) (Giffin, 2008; 2009). Using Lagrange multipliers, α and β, one can write

$$\frac{1}{N}\ln W = -\sum_i^m f_i \ln f_i - \alpha\left(\sum_{i=1}^{m} f_i - 1\right) - \beta\left(\sum_i f_i \epsilon_i - E\right) \tag{36}$$

and maximize $(\ln W)/N$ as

$$\begin{aligned} \Delta\left(\frac{1}{N}\ln W\right) &= -\sum_i \Delta f_i \ln f_i - \sum_i f_i \frac{\Delta f_i}{f_i} - \alpha\sum_i \Delta f_i - \beta\sum_i (\Delta f_i)\,\epsilon_i \\ &= \sum_i \Delta f_i\,(-\ln f_i - 1 - \alpha - \beta\epsilon_i) \\ &= 0 \end{aligned} \tag{37}$$

Therefore, f_i is calculated as

$$f_i = e^{-(1+\alpha+\beta\epsilon_i)} \tag{38}$$

Eq. (30) indicates the sum of f_i should be one:

$$1 = \sum_i f_i = e^{-1-\alpha} \sum_i e^{-\beta\epsilon_i} \tag{39}$$

so that

$$e^{-1-\alpha} = \left[\sum_i e^{-\beta\epsilon_i}\right]^{-1} = \frac{1}{Z} \tag{40}$$

where Z is partition function, defined as

$$Z = \sum_i e^{-\beta\epsilon_i} \tag{41}$$

The final form of function f_i is written as

$$f_i = \frac{e^{-\beta\epsilon_i}}{Z} \tag{42}$$

which makes the mean energy E represented in terms of the partition function:

$$E = \sum_i f_i\epsilon_i = \frac{1}{Z}\sum_i \epsilon_i e^{-\beta\epsilon_i} = -\frac{1}{Z}\frac{\partial}{\partial\beta}\left(\sum_i e^{-\beta\epsilon_i}\right) = -\frac{\partial \ln Z}{\partial\beta} \tag{43}$$

1.2.3 Canonical ensemble

In classical thermodynamics, there are seven primary quantities: (1) the number of particles (or molecules) N, (2) the volume V of the system containing the particles, (3) the temperature T, (4) the pressure P due to collisions of particles on box walls, (5) the total energy E, (6) the entropy S measuring the disorderness of the system, and (7) the chemical potential of μ (i.e., molar Gibbs free energy). An ensemble sets three (out of seven) variables to constants and defines a characteristic energy-function (i.e., a thermodynamics function that has a unit of energy) with the three constant variables as arguments. The partition function determines the characteristic energy-function. The other four variables are determined using the energy function and its partial derivatives with respect to the three variables chosen for the ensemble. For example, the canonical ensemble sets N, V, and T as constants and defines the Helmholtz free energy as

$$F = E - TS = -k_B T \ln Z \tag{44}$$

where k_B is the Boltzman constant and Z is the (canonical) partition function:

$$Z\left(N, V, T\right) = \frac{1}{N! h^{3N}} \int e^{-H(\boldsymbol{\Gamma})/k_B T} d\boldsymbol{\Gamma} \tag{45}$$

where h is Planck's constant, $H(\boldsymbol{\Gamma})$ is the Hamiltonian, $\boldsymbol{\Gamma}$ and $\mathrm{d}\boldsymbol{\Gamma} = \mathrm{d}^N\boldsymbol{r}\mathrm{d}^N\boldsymbol{p}$ represent a specific state and the infinitesimal element, respectively, in the phase space of $3N \times 3N$ dimension. Given a specific Hamiltonian as the sum of kinetic and potential energies of N

particles in a conservative field, i.e.,

$$H = \sum_{i=1}^{N} \left(\frac{\overrightarrow{P_i} \cdot \overrightarrow{P_i}}{2m_i} + V\left(\overrightarrow{r_i}\right) \right) \tag{46}$$

and using the infinitesimal relationship of the Helmholtz free energy

$$dF = \mu dN - PdV - SdT \tag{47}$$

the rest of four variables in the canonical ensemble are calculated as

$$\mu = \left(\frac{\partial F}{\partial N} \right)_{V,T} \tag{48}$$

$$P = -\left(\frac{\partial F}{\partial V} \right)_{N,T} \tag{49}$$

$$S = -\left(\frac{\partial F}{\partial T} \right)_{N,V} \tag{50}$$

$$E = -k_B T \ln Z - TS \tag{51}$$

where subscripts N, V, and T of the partial differentials are kept constant.
On the other hand, one can write mean energy E as an ensemble average of the Hamiltonian:

$$E = \langle H \rangle = \int H \cdot \rho(\Gamma)\, d\Gamma \tag{52}$$

where ρ is the probability density function (PDF) that quantifies the chances of the system having a specific value of the Hamiltonian:

$$\rho = \frac{e^{-\beta H}}{Z} \tag{53}$$

where $\beta = 1/k_B T$. Using the PDF, the second law of thermodynamics can be revisited by calculating

$$\Delta E = \int \Delta H \cdot \rho \, d\Gamma + \int H \cdot \Delta\rho \, d\Gamma \tag{54}$$

Here, the first term of Eq. (54) indicates the average of the Hamiltonian charge, which must be heat provided to the system by the surroundings:

$$\int \Delta H \cdot \rho \, d\Gamma = \delta Q \tag{55}$$

where "δ" has a similar meaning to đ, implying an inexact differential of which the associated quantity (e.g., Q) is a path function. The second term of Eq. (54) indicates energy change in part by variation of probability distributions, $\Delta\rho$. As a response to the applied (infinitesimal) heat δQ, the system does some work without noticeable change of the total energy; and their difference is stored as the internal energy change. In a reversible process, this work accompanies an infinitesimal volume expansion keeping the system pressure invariant. As the system is heated by the surroundings, the molecules will have more thermodynamic states because the entropy never decreases. Then, the change in the probability distribution is negative and the second integral term of Eq. (54) is expressed as negative (infinitesimal) work:

$$\int H \cdot \Delta\rho \, \mathrm{d}\Gamma = -\delta W \tag{56}$$

By combining Eqs. (54) –(56), we revisited the second law of the thermodynamics from the view point of statistical mechanics, i.e.,

$$\Delta E = \delta Q - \delta W \tag{57}$$

which is equivalent to Eq. (25).

2. Monte Carlo sampling

2.1 Importance sampling

As the first law of thermodynamics describes energy changes as heat is absorbed or subtracted by work done, the average energy of canonical ensemble of subsystems in an equilibrium state is described as its ensemble average (Eq. 52). Constraints embedded in the canonical ensemble are N, V, and T. The other four thermodynamic variables such as E, P, μ, and S are calculated using the definition of ensemble average:

$$\langle \mathcal{A} \rangle = \int \mathcal{A} \cdot \rho(\mathbf{\Gamma}) \, d\mathbf{\Gamma} \tag{58}$$

where $\mathcal{A}$ is an thermodynamic variable of interest. Numerical integration of $\mathcal{A} \cdot \rho(\mathbf{\Gamma})$ in phases space is almost formidable due to the large dimension of $3N \times 3N$. Another approach is to generate a number of samples of $\mathcal{A}$ where $\rho(\mathbf{\Gamma})$ is significant.

For example, if a Hawaii researcher tries to build a PDF of shoppers at Ala Moana Shopping Center[1], then the researcher needs to have a number of assistants who stay for a given time (e.g., 10 minutes) at uniformly distributed stations in the mall areas and count the number of people passing by them. This is a very direct way to build the shopper PDF throughout the mall. If the researcher wants to know the average number of shoppers per 10 minutes, then he/she will multiply the number of shoppers counted at each station to the normalized frequency of shoppers at each station and sum over all the stations. For simplicity, assume that a shopper is not simultaneously double counted by assistants at different stations, and all shoppers are counted. Using this method, construction of a master PDF is a primary, time-consuming step before calculating any statistical quantity of interest. Isn't there any faster method to accurately estimate the mean number of shoppers? Assume the researcher knows, by previous knowledge, the shopper PDF in the mall. Then, the survey stations can be distributed following the shopper PDF. Once each station provides the number of shoppers counted during 10 minutes, then the sum of the numbers of shoppers of all stations can be a good estimation of the mean shopper number during 10 minutes!

Monte Carlo sampling, called importance sampling, follows the same idea. Instead of spanning all the possible cases and measuring a quantity of interest, sampling can be done following a PDF of thermodynamic states. Obviously, this approach to use the PDF for sampling, which should be obtained by sampling, is controversial. The Hawaii researcher can use a dynamic sampling method to reduce the sampling number required to build a trustful PDF. Initially, the survey stations can be distributed with a uniform, random, or mixed manner throughout the mall. A station assigned to Leia[2] is located between a parking lot and

---

[1] The largest shopping mall in Honolulu, Hawaii, the fifteenth largest shopping mall in the United States, and the largest open-air shopping center in the world

[2] Hawaiian female name, meaning child of heaven

a shopping area. Leia counted the number of people, say s_1 at position $\boldsymbol{x}_1$. She randomly selected a new position $\boldsymbol{x}_2$ (not far away from her original location $\boldsymbol{x}_1$), counted the number of shoppers, say s_2, and found that $s_2 > s_1$. Then, she updated her numbers from s_1 to s_2. Her third position $\boldsymbol{x}_3$ is chosen randomly near $\boldsymbol{x}_2$. At this time, s_3 is less than s_2. Now, she thinks that if she keeps selecting new positions with higher numbers than that at a previous station, the measurement is not correct and statistics will overestimate the total number of people in the mall. In this case, she needs a criterion to accept movement to a station where the number of people counted will be lower.

She decides on a simple protocol as follows:

- If a new position provides more people, then stay at the new position and update the number.
- If the number of shoppers decreases by moving to a new position, she stay at the new position if she tosses heads on a coin.

By doing this, she can measure more frequently where there are more shoppers, but not completely discard data from stations having less shoppers. She can test various criteria to stay at a new position with a lower number of shoppers. If the ratio of the number of shoppers at the new station to that of old, i.e, $N_{\text{new}}/N_{\text{old}}$ is less than 0.5, she returns to the old position, discarding the number counted at the new position. Because the new position is rejected, she counts the number of people at the old position one more time before she randomly selects a new position. Including the random selection of $\boldsymbol{x}_1$, the number of positions used for measurement should be equal to the number of accepted measurements. Some of positions must be used multiple times. After a certain number of measurements, let's say 50, Leia can assemble a series with the number of shoppers, i.e., $\{N_1, N_2, N_3, \ldots, N_i, \ldots, N_{50}\}$. The difference between N_{i+1} and N_i must be large if i is small. Leia can take an average of the number of people using only the later half of the data set, i.e, $i = 26 - 50$. If the researcher has more assistants, he or she can sum the number of shoppers counted by the assistants, divide it by the number of assistants, and calculate the average number of people seen per 10 minutes at the shopping mall.

2.2 Formalism

Monte Carlo simulation is in general an integration process in a large dimensional space. As stated above, the key issue is to reduce the sampling numbers using the importance sampling technique, i.e., sampling more data where the probability density is higher. The sampling frequency distribution should resemble the PDF of the thermodynamic states. Monte Carlo simulation starts at a location in the phase space. The next position is selected to satisfy that a series of accepted locations follow the frequency distribution of the thermodynamic state. Therefore, searching probable states in the phase space is a self-consistent process that supports the PDF.

After integration with respect to N-particle momentum, $\mathrm{d}\boldsymbol{p}^N$, the probability distribution function of a particle configuration state $\boldsymbol{R}\,(=\{\boldsymbol{r}_1, \boldsymbol{r}_2, \cdots, \boldsymbol{r}_i, \cdots, \boldsymbol{r}_N\})$ occurring is

$$\rho(\boldsymbol{R}) = \frac{e^{-\beta V(\boldsymbol{R})}}{Z_R} \tag{59}$$

where

$$Z_R = \int e^{-\beta V(\boldsymbol{R})} \mathrm{d}\boldsymbol{R} \tag{60}$$

Now, we define $W(\boldsymbol{R}'|\boldsymbol{R})$ as the probability to make a transition from state $\boldsymbol{R}$ to $\boldsymbol{R}'$, i.e., the conditional probability of occurring state $\boldsymbol{R}$ given that state $\boldsymbol{R}'$ has occured. Two constraints $W(\boldsymbol{R}'|\boldsymbol{R})$ are:

(a) sum rule

$$\int W(\boldsymbol{R}'|\boldsymbol{R})\mathrm{d}\boldsymbol{R}' = 1 \tag{61}$$

(b) detailed balance

$$W(\boldsymbol{R}'|\boldsymbol{R})\,\rho(\boldsymbol{R}) = W(\boldsymbol{R}|\boldsymbol{R}')\,\rho(\boldsymbol{R}') \tag{62}$$

In Eq. (62), left-hand-side is the probability of state $\boldsymbol{R}$ occurring multiplied by the transition probability to state $\boldsymbol{R}'$ given $\boldsymbol{R}$; and right-hand-side is the probability of state $\boldsymbol{R}'$ occurring multiplied by the transition probability to state $\boldsymbol{R}$ given $\boldsymbol{R}'$. Therefore, each represents the probability of both configuration states $\boldsymbol{R}$ and $\boldsymbol{R}'$ occurring in canonical ensemble, i.e., $P(\boldsymbol{R} \cap \boldsymbol{R}')$.

The transition probability W is described as

$$W(\boldsymbol{R}'|\boldsymbol{R}) = A(\boldsymbol{R}'|\boldsymbol{R})\,T(\boldsymbol{R}'|\boldsymbol{R}) \tag{63}$$

where $A(\boldsymbol{R}'|\boldsymbol{R})$ is the acceptance probability of the move from $\boldsymbol{R}$ to $\boldsymbol{R}'$ and $T(\boldsymbol{R}'|\boldsymbol{R})$ is the selection probability of a new configuration $\boldsymbol{R}'$ given an old $\boldsymbol{R}$, chosen to satisfy

$$\int T(\boldsymbol{R}'|\boldsymbol{R})\mathrm{d}\boldsymbol{R}' = 1 \tag{64}$$

such that $\boldsymbol{R}$ and $\boldsymbol{R}'$ are accessible to each other.

If configuration $\boldsymbol{R}'$ is a higher energy state than $\boldsymbol{R}$, which indicates that a transition from state $\boldsymbol{R}'$ to $\boldsymbol{R}$ is always accepted, i.e.,

$$A(\boldsymbol{R}|\boldsymbol{R}') = 1 \tag{65}$$

Substitution of Eq. (63) into Eq. (62) yields

$$A(\boldsymbol{R}'|\boldsymbol{R}) = \frac{T(\boldsymbol{R}|\boldsymbol{R}')\,\rho(\boldsymbol{R}')}{T(\boldsymbol{R}'|\boldsymbol{R})\,\rho(\boldsymbol{R})} \tag{66}$$

where the form of $T(\boldsymbol{R}'|\boldsymbol{R})$ can be arbitrarily chosen. Now, we define

$$q(\boldsymbol{R}'|\boldsymbol{R}) = \frac{T(\boldsymbol{R}|\boldsymbol{R}')}{T(\boldsymbol{R}'|\boldsymbol{R})} \exp\left(-\beta\left[V(\boldsymbol{R}') - V(\boldsymbol{R})\right]\right) \tag{67}$$

and represent a general form of the acceptance probability:

$$A(\boldsymbol{R}'|\boldsymbol{R}) = \min\left[1, q\left(\boldsymbol{R}'|\boldsymbol{R}\right)\right] \tag{68}$$

If $V(\boldsymbol{R}') < V(\boldsymbol{R})$, $q(\boldsymbol{R}'|\boldsymbol{R}) > 1$ and therefore $A(\boldsymbol{R}'|\boldsymbol{R}) = 1$. Otherwise, $A(\boldsymbol{R}'|\boldsymbol{R}) = q(\boldsymbol{R}'|\boldsymbol{R}) < 1$.

2.3 Metropolis Monte Carlo (MMC)

In principle, it is possible to make simultaneous N-particle moves. However, in practice, a one particle move is preferred because the acceptance probability of the N-particle move is exceedingly small for a reasonable step size. For this reason, we restrict ourselves to the one-particle move at a Monte Carlo step, and simplify Eq. (68) to

$$A(\boldsymbol{r}_i'|\boldsymbol{r}_i) = \min\left[1, q\left(\boldsymbol{r}_i'|\boldsymbol{r}_i\right)\right] \tag{69}$$

which implies that, during the transition of particle i from $\boldsymbol{r}_i$ to $\boldsymbol{r}_i'$, all other particles will stay at their original positions $\boldsymbol{r}_j$ for $j = 1, 2, \cdots, i-1, i+1, \cdots, N$.

To implement the Metropolis solution of the acceptance probability, the selection probability needs to be specified, which is designed to move particle i from $\boldsymbol{r}_i$ into any one of its neighboring position $\boldsymbol{r}_i'$ in a cubic region $\mathscr{R}$ with an equal probability. See Fig. 3 for details. The cube is centered at $\boldsymbol{r}_i$ and has sides of $2\delta r_{\max}$. Then, the random displacement $\Delta\boldsymbol{r}_i = \boldsymbol{r}_i' - \boldsymbol{r}_i$ is selected via

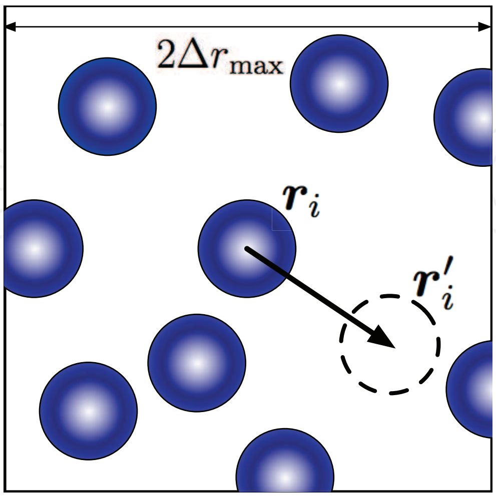


Fig. 3. Trial movement of particle i from position $\boldsymbol{r}_i$ to $\boldsymbol{r}_i'$ that can be any point within the square domain $\mathscr{R}$ of each side $2\Delta r_{\max}$.

$$\Delta x_i = x_i' - x_i = (2\xi_1 - 1.0) \times \Delta r_{\max} \tag{70a}$$
$$\Delta y_i = y_i' - y_i = (2\xi_2 - 1.0) \times \Delta r_{\max} \tag{70b}$$
$$\Delta z_i = z_i' - z_i = (2\xi_3 - 1.0) \times \Delta r_{\max} \tag{70c}$$

where ξ_1, ξ_2, and ξ_3 are independent uniform random numbers between 0 and 1. The value of $\Delta r_{\max}$ must be optimally chosen. If $\Delta r_{\max}$ is too large, then the acceptance probability can be small due to physical overlap between rigid particles and/or the energy difference can be much larger than $k_B T$. The selection probability of $\boldsymbol{r}_i'$ from $\boldsymbol{r}_i$ is then

$$T(\boldsymbol{r}_i'|\boldsymbol{r}_i) = \begin{cases} \tau_0 & \boldsymbol{r}_i' \in \mathscr{R} \\ 0 & \boldsymbol{r}_i' \notin \mathscr{R} \end{cases} \tag{71}$$

where τ_0 is a constant, which can be chosen as $1/\left(2\Delta r_{\max}\right)^3$ or $1/N_{\mathscr{R}}$ where $N_{\mathscr{R}}$ is the finite number of a new position. The numerical value of τ_0 does not influence the evaluation of the acceptance probability because the constant character of τ_0 implies its intrinsic symmetricity with respect to the sequence of $\boldsymbol{r}_i$ and $\boldsymbol{r}_i'$, which provides the final form of

$$q(\boldsymbol{r}_i'|\boldsymbol{r}_i) = \frac{\tau_0}{\tau_0}\frac{\rho(\boldsymbol{r}_i')}{\rho(\boldsymbol{r}_i)} = e^{-\beta[V(\boldsymbol{r}_i') - V(\boldsymbol{r}_i)]} \tag{72}$$

from Eq. (67) and further the acceptance probability of the one-particle move as

$$A(\boldsymbol{r}_i'|\boldsymbol{r}_i) = \min\left[1, q\left(\boldsymbol{r}_i'|\boldsymbol{r}_i\right)\right] \tag{73}$$

as suggested by Metropolis et al. (1953).

2.4 Force Bias Monte Carlo (FBMC)

Pangali et al. (1978) proposed the modified selection probability along the direction of forces. The primary reason is that particle movements in (deterministic) molecular dynamics per time step is usually biased in the direction of the intermolecular forces and torques, whereas the moves sampled according to the standard Metropolis algorithm are chosen randomly in an arbitrary direction. Therefore, if the influence of forces acting on a particle can be included in selecting a new position, then the Monte Carlo can be more accurate, potentially avoiding congestion near a bottleneck in phase space.

Expansion of the potential function $V(\boldsymbol{r}_i)$ around $\boldsymbol{r}_i$ gives for the one-particle move

$$T(\boldsymbol{r}_i'|\boldsymbol{r}_i) = \begin{cases} C^{-1}\exp\left[\beta\lambda \boldsymbol{F}_i(\boldsymbol{r}_i)\cdot\Delta\boldsymbol{r}_i\right] & \boldsymbol{r}_i' \in \mathscr{R} \\ 0 & \boldsymbol{r}_i' \notin \mathscr{R} \end{cases} \tag{74}$$

where λ is an arbitrary parameter and $\boldsymbol{F}_i(\boldsymbol{r}_i) = -\nabla_{\boldsymbol{r}_i}V(\boldsymbol{r}_i)$ is a force vector acting on particle i from all other particles, and C is a normalization constant that depends on λ and $\boldsymbol{F}_i$. The normalization condition of Eq. (64) gives

$$C = \int\int\int_{\boldsymbol{r}-\Delta\boldsymbol{r}_{\max}}^{\boldsymbol{r}+\Delta\boldsymbol{r}_{\max}} \mathrm{d}x'\mathrm{d}y'\mathrm{d}z'\ \exp\left[\beta\lambda\left\{F_x\cdot(x'-x)+F_y\cdot(y'-y)+F_z\cdot(z'-z)\right\}\right] \tag{75}$$

where the particle index i is omitted since it does not change the final representation of C. The integration with respect to x' calculates

$$\int_{x-\Delta r_{\max}}^{x+\Delta r_{\max}} \mathrm{d}x' \exp\left[\beta\lambda F_x\cdot(x'-x)\right] = \frac{2\sinh\left[\beta\lambda F_x \Delta r_{\max}\right]}{\beta\lambda F_x}$$

and so

$$C = \frac{8\sinh\left[\beta\lambda F_x\Delta r_{\max}\right]\sinh\left[\beta\lambda F_y\Delta r_{\max}\right]\sinh\left[\beta\lambda F_z\Delta r_{\max}\right]}{(\beta\lambda)^3 F_x F_y F_z} \tag{76}$$

Note that F_x, F_y, and F_z are force components that particle i experiences at $\boldsymbol{r} = (x,y,z)$. The coefficient C' of the selection probability of $\boldsymbol{r}_i$ from $\boldsymbol{r}_i'$ is accordingly calculated as

$$C' = \frac{8\sinh\left[\beta\lambda F_x'\Delta r_{\max}\right]\sinh\left[\beta\lambda F_y'\Delta r_{\max}\right]\sinh\left[\beta\lambda F_z'\Delta r_{\max}\right]}{(\beta\lambda)^3 F_x' F_y' F_z'} \tag{77}$$

with F_x', F_y', and F_z' at $\boldsymbol{r}' = (x',y',z')$. So that

$$\begin{aligned} q(\boldsymbol{r}'|\boldsymbol{r}) &= \frac{T(\boldsymbol{r}|\boldsymbol{r}')\rho(\boldsymbol{r}')}{T(\boldsymbol{r}'|\boldsymbol{r})\rho(\boldsymbol{r})} \\ &= \frac{C'}{C}\exp\left(-\beta\left[V(\boldsymbol{r}')-V(\boldsymbol{r})+\lambda(\boldsymbol{F}+\boldsymbol{F}')\cdot\Delta\boldsymbol{r}\right]\right) \\ &= \exp\left(-\beta\left[\delta V+\lambda(\boldsymbol{F}+\boldsymbol{F}')\cdot\Delta\boldsymbol{r}+\Delta\mathcal{W}_{\mathrm{FB}}\right]\right) \end{aligned} \tag{78}$$

where

$$\Delta V = V(\boldsymbol{r}') - V(\boldsymbol{r}) \tag{79}$$

and

$$\Delta\mathcal{W}_{\mathrm{FB}} = k_B T \ln\left(\frac{C'}{C}\right) \tag{80}$$

If the maximum displacement $\Delta r_{\max}$ is set to be small, then C'/C can be approximated using the Taylor expansion. If we consider terms only in the x-coordinate,

$$\left(\frac{C'}{C}\right)_x = \frac{\sinh[\eta + \Delta\eta]}{\eta + \Delta\eta} \frac{\eta}{\sinh[\eta]} \tag{81}$$

where $\eta = \beta\lambda F_x \Delta r_{\max}$ and $\Delta\eta = \beta\lambda\Delta F_x \Delta r_{\max}$. Using

$$\frac{\sinh t}{t} = 1 + \frac{1}{6}t^2 + \cdots \tag{82}$$

one approximates C'/C as

$$\begin{aligned}\left(\frac{C'}{C}\right)_x &\approx \frac{1 + \frac{1}{6}(\eta + \Delta\eta)^2}{1 + \frac{1}{6}\eta^2} \\ &= \left[1 + \frac{1}{6}(\eta + \Delta\eta)^2\right]\left[1 - \frac{1}{6}\eta^2 + O\left(\eta^4\right)\right] \\ &= 1 + \frac{1}{6}\left[2\eta\Delta\eta + (\Delta\eta)^2\right] + O\left(\eta^4\right) \\ &\approx 1 + \frac{1}{6}\left[2F_x\Delta F_x + \Delta F_x^2\right]\beta^2\lambda^2\Delta r_{\max}^2 \\ &\approx \exp\left(\frac{1}{6}\left[2F_x\Delta F_x + \Delta F_x^2\right]\beta^2\lambda^2\Delta r_{\max}^2\right) \end{aligned} \tag{83}$$

so that in general

$$\Delta\mathcal{W}_{\mathrm{FB}} \approx \frac{1}{6}\left(2\boldsymbol{F}\cdot\Delta\boldsymbol{F} + \Delta\boldsymbol{F}\cdot\Delta\boldsymbol{F}\right)\beta\lambda^2\Delta r_{\max}^2 \tag{84}$$

Finally, a move from $\boldsymbol{r}$ to $\boldsymbol{r}'$ of i particle is accepted with the probability:

$$A(\boldsymbol{r}'|\boldsymbol{r}) = \min\left[1, q\left(\boldsymbol{r}'|\boldsymbol{r}\right)\right] \tag{85}$$

where

$$q(\boldsymbol{r}'|\boldsymbol{r}) = \exp\left(-\beta\left[\delta V + \lambda(2\boldsymbol{F} + \Delta\boldsymbol{F})\cdot\Delta\boldsymbol{r} + \frac{1}{6}\beta\lambda^2\Delta r_{\max}^2(2\boldsymbol{F} + \Delta\boldsymbol{F})\cdot\Delta\boldsymbol{F}\right]\right) \tag{86}$$

An alternative representation with $\lambda = 1/2$ is

$$q(\boldsymbol{r}'|\boldsymbol{r}) = \exp\left(-\beta\left[\delta V + \langle\boldsymbol{F}\rangle\cdot\Delta\boldsymbol{r}_{\mathrm{FB}}\right]\right) \tag{87}$$

where $\langle\boldsymbol{F}\rangle$ indicates the average forces between position $\boldsymbol{r}$ and $\boldsymbol{r}'$:

$$\langle\boldsymbol{F}\rangle = \frac{1}{2}\left(\boldsymbol{F} + \boldsymbol{F}'\right) = \boldsymbol{F} + \frac{1}{2}\Delta\boldsymbol{F} \tag{88}$$

and $\Delta\boldsymbol{r}_{\mathrm{FB}}$ is the proposed displacement for a move:

$$\Delta\boldsymbol{r}_{\mathrm{FB}} = \Delta\boldsymbol{r} + \frac{1}{12}\beta\Delta r_{\max}^2\Delta\boldsymbol{F} \tag{89}$$

The first and second terms in the RHS of Eq. (89) are under the influence of random forces (from surrounding solvent molecules) and systematic forces (from other nearby Brownian particles), respectively. Note that $\beta^{-1}|\Delta\boldsymbol{F}|\Delta r_{\max}$ has an energy unit. The simplified representation of the acceptance probability q of Eq. (87) evolves from two conditions: (1) $\lambda = 1/2$, and (2) $\Delta r_{\max}$ is small enough (typically in comparison to particle size) to ensure that $\beta\lambda|\boldsymbol{F}|\Delta r_{\max} \ll 1$. With the same conditions, the "Smart Monte Carlo (SMC)" provides identical representations of Eqs. (87) - (89) developed by Rossky et al. (1978). Northrup and McCammon (1980) have used SMC to investigate protein structure fluctuations.

In the FB algorithm, $\Delta\boldsymbol{r}$ is randomly chosen using a uniform random number generator as indicated in Eq. (70): however, in the SMC algorithm, $\Delta\boldsymbol{r}/\Delta r_{\max}$ is chosen to have zero mean and unit variance; in other words,

$$\langle \Delta\boldsymbol{r} \rangle = 0 \tag{90a}$$

$$\langle \Delta\boldsymbol{r} \cdot \Delta\boldsymbol{r} \rangle = \Delta r_{\max}^2 \tag{90b}$$

and the isotropy indicates

$$\langle \Delta x \rangle = \langle \Delta y \rangle = \langle \Delta z \rangle = 0 \tag{91a}$$

$$\langle (\Delta x)^2 \rangle = \langle (\Delta y)^2 \rangle = \langle (\Delta z)^2 \rangle = \frac{1}{3}\Delta r_{\max}^2 \tag{91b}$$

A Gaussian random number with zero mean and unit variance can be generated using two typical methods.

A. Box and Muller's Algorithm

Box and Muller (1958) Let U_1 and U_2 be independent random variables from the same rectangular density function on the interval $(0, 1)$. Consider the random variables:

$$X_1 = \sqrt{-2\ln U_1}\,\cos 2\pi U_2 \tag{92}$$

$$X_2 = \sqrt{-2\ln U_1}\,\sin 2\pi U_2 \tag{93}$$

Then, $-\infty \le X_1, X_2 \le +\infty$ will be a pair of independent random variables from the same normal distribution with mean zero, and unit variance. Either X_1 or X_2 can be used if only one Gaussian random number is necessary at a time.

B. Summation Algorithm

Allen and Tildesley (1987) This method involves two steps and the generation of 12 uniform random variates:

1. generate 12 uniform random variables, $U_1, \cdots, U_{12}$ in range $(0, 1)$;
2. calculate $X = \sum_{i=1}^{12} U_i - 6$

This method yields number X's which are sampled from an approximately normal distribution (by virtue of the central limit theorem of probability). Clearly, random variates outside the range $(-6, +6)$ will never be generated using this method, but it is adequate for most Monte Carlo purposes, and is quite fast.

2.5 Move Bias Monte Carlo (MBMC)

In natural and engineered systems, molecules and particles are considered as point masses and spherical objects, respectively. However, long polymer chains with finite length are of great importance in chemical engineering processes. Since a polymer can be viewed as a linear connection of many identical monomers, trial movement of a polymer requires constraints such as sequence, bond length, and bond angle of associated monomers.

Mapping the movement of a polymer chain is an interesting problem. de Gennes (1971) developed the reptation model where a polymer chain of n monomers moves like a reptile. When the position of the first monomer is randomly selected and accepted without any overlap between the other monomers, the second monomer moves to the previous position of the first, the third monomer moves to the previous position of the second monomer, and so forth: simply, $\boldsymbol{r}_1^{\mathrm{old}} \to \boldsymbol{r}_1^{\mathrm{new}}$ and $\boldsymbol{r}_i \to \boldsymbol{r}_{i-1}$ for $i = 2 - n$. In programming, this is actually equivalent to moving the last monomer to the position of the first monomer and updating the monomer index such that

$$i \to \mathrm{mod}(i, m) + 1 \tag{94}$$

i.e., $1 \to 2$, $2 \to 3$, ..., $i \to (i+1)$, $(m-1) \to m$, and $m \to 1$.

This reptation model is efficient when polymers are in a dense system so that their lateral movements are restricted by volume exclusion. However, when polymer rheology is of interest, one can use an isothermal-isobaric ensemble with a constant number of particles (i.e., monomers in polymer solution), pressure, and temperature. Simulation starts with a sparse distribution of polymers of random shapes in an arbitrarily chosen large volume. Configurations of polymers and system volume change during the NPT simulation to reach an equilibrium state that calculates a mean pressure close to the preset value of the pressure. While the volume shrinks, polymers find energetically nested positions. In this case, an effective sampling in the phase space should include lateral movement of monomers within a polymer in addition to the reptation. To avoid configurational jamming of polymers in a canonical ensemble, one can use parallel tempering which simulates M replicas of the original system of interests at different temperatures and swap a pair of configurations comparing energy between the two subsystems (Earl and Deem, 2005).

A partial polymer can have a new configuration. A terminal section of a chain, let't say m monomers, is annihilated and then regrown one by one. Each of the new m monomers can be created at a random position near the previous one in the sequence. The energy difference between the new and old terminal sections is calculated, and the standard Metropolis Monte Carlo scheme is used to accept or reject the new configuration of the terminal section. This is called configurational bias Monte Carlo (CBMC), originally proposed by Rosenbluth and Rosenbluth (1955) and further developed by Siepmann and Frenkel (1992) and de Pablo et al. (1992). As the force bias MC was designed to simulate water and aqueous solutions (Pangali et al., 1978), it is possible to combine the move bias and force bias Monte Carlo algorithms for an efficient sampling of polymer configurations in phase space.

3. Hybrid Monte Carlo

The term "hybrid" indicates a combination of two different elements. In general, thermodynamic states can be classified into equilibrium and non-equilibrium states, often called "static" and "dynamic", respectively. The true equilibrium state exists in an isolated system through which neither mass nor energy can transfer. Therefore, no thermodynamic quantity in an isolated system changes with respect to time, which is mathematically depicted as $\partial[\,]/\partial t = 0$. In engineering processes, mass and energy continuously enter and leave from an operational unit. When the mass/energy transport rate is kept constant, the system is said to be in a steady state or dynamic equilibrium. Principles and theories of statistical mechanics developed under pure equilibrium are often adopted to explain and analyze engineering processes at the steady state.

Particles found in gaseous and liquid flows are influenced by external force fields and inter-particle interactions. Hydrodynamic drag forces exerted on suspended particles in a fluid medium contribute to convective (or advective) transport of particles in a plethora of natural and engineering processes. Brownian motion (Brown, 1828) of particles involves a tremendous amount of bombastic collisions of fluid molecules with suspended particles. Relative motion of particles flowing in fluid media generate resistant forces in the opposite direction to the particle's translational motion. To rigorously deal with this phenomena, Newton's second law should be numerically solved for positions and velocities of particles as a function of time. The exerted forces can be classified into conservative and dissipative ones; the former include external and interparticle forces solely depending on particle positions, and the latter on hydrodynamic drag force proportional to particle velocities. Because the Monte Carlo methods described thus far propose efficient ways of sampling in phase space, it should be ensured that the system to be investigated is in a static equilibrium state of conservative force field. Then, isn't there any method to include the dissipative force with the standard Monte Carlo method and phenomenologically simulate steady-state configurations of many particles in a fluid system? The answer is the hybrid Monte Carlo.

Arya and Panagiotopoulos (2004) used a coarse-grained lattice Monte Carlo approach to model surfactants, while the shear is implemented by assigning a "pseudopotential" as included into the MC acceptance criteria. Their model systems consist of cylindrical micelles located between two impermeable walls, and each micelle is depicted as 4 head and 4 tail monomers, i.e., 4H4T lattice surfactants. For the sake of simplicity, they impose attractive potential energy between only two tail monomers associated in the same or different micelles: $\epsilon_{\mathrm{TT}}/k_B T = -2$, and $\epsilon_{\mathrm{TH}} = \epsilon_{\mathrm{HH}} = 0$ where subscripts H and T indicate head and tail, respectively. A fictitious potential energy gradient ∇U_d to the drag force experience by each monomer is given by

$$\nabla U_d = \zeta v_x \tag{95}$$

where v_x is the fluid velocity in the x direction and ζ is a friction coefficient. In a linear shear flow, $v_x = \dot{\gamma} y$ where $\dot{\gamma}$ is the shear rate and y is the coordinate perpendicular to the slit. The pseudo-energy difference term of monomer i moving from (x_1, y_1, z_1) to (x_2, y_2, z_2) is given by

$$\nabla U_{d,i} = -\zeta \dot{\gamma} \langle y \rangle \left(x_2 - x_1 \right) \tag{96}$$

where $\langle y \rangle = \left(y_1 + y_2 \right) / 2$. The final acceptance criterion for an MC move of an amphiphile (4H4T monomers) is assigned as

$$P_{\mathrm{acc}} = \min \left(1, \exp \left[- \left(\Delta V + \sum_{i=1}^{N=8} \Delta U_{d,i} \right) / k_B T \right] \right) \tag{97}$$

where ΔU is the usual interaction energy contribution, and the shear contribution has been summed over the 8 monomers per amphiphile. Three dimensional lattice MC simulations showed various alignments due to the shear rate and dimensionless temperature (defined as $k_B T/\epsilon_{\mathrm{TT}}$). The shear force breaks micelles perpendicular to the shearing direction into small ones, and these smaller micelles align and grow in the shearing direction. This approach can be considered as a hybrid MC which includes dissipative force as an origin of the pseudo-potential energy.

Kim et al. (2001) already used a similar approach of combining conservative inter-particle potential energy and dissipative hydrodynamic drag force in a hybrid Monte Carlo to study shear-induced micellar deformation. Lennard-Jones (LJ) potentials were assumed for TT and TH pairs with different potential well depths, i.e., ϵ_{TT} and ϵ_{TH}. The HH pair interaction include LJ as well as Coulomb potentials to characterize electrostatic repulsion between two head groups. The shear force exerted on each monomer is calculated as

$$F_{x,j}(y) = 3\pi\mu\sigma_{jj}\dot{\gamma} y_j \tag{98}$$

where j is H or T, σ is diameter, and μ is the solvent viscosity. Shear force is treated as a bias influence in a trail MC move of a tail or head. The standard force bias MC was employed by switching the gradient of the pair potential energy with the hydrodynamic force, and a trial move is accepted following the acceptance probability of Eq. (87). Micellar deformation was investigated as a function of Peclet number (Pe) and rapid transition from spherical to oval shapes was observed near $\mathrm{Pe} = 1$. This indicated that if the move displacement is small enough, then Brownian dynamics, smart MC, and force bias MC are equivalent (Chen and Kim, 2004).

Kim et al. (2001) and Arya and Panagiotopoulos (2004) used similar approaches of including hydrodynamic force in the Metropolis MC. One important aspect which was ignored by both papers was that the microscopic hydrodynamic force is significantly influenced by the number of neighbors and their relative positions. When particles accumulate and form a porous media, drag force increases with respect to the particle volume fraction. Then, Stokes' drag needs to be modified quantitatively. Happel (1958) developed a tangential-stress-free cell model, which maps a uniform distribution of particles to a sphere embedded in a concentric spherical cell. On the surface of the cell, tangential stress was assumed to be zero. Then, the ratio of force exerted on a spherical particle of radius a within a sphere-packed porous media of volume fraction ϕ to that on an isolated sphere is given as

$$\Omega = \frac{F(\phi)}{6\pi\mu a u} = \frac{6 - 9\phi^{1/3} + 9\phi^{5/3} - 6\phi^2}{6 + 4\phi^{5/3}} \tag{99}$$

where u is the superficial velocity. To calculate the force ratio Ω, the local volume fraction near a particle that will move should be calculated. Because the potential energy is assumed to be pairwise, center-to-center distances between a particle and all others need to be calculated. One can make an equivalent cell whose volume is equal to that of the search domain of new position, i.e., $(2\Delta r_{\max})^3$ of Fig. 3, count the number of particles within the cell, and calculate the local volume fraction near the particle that is about to move. This technique was developed to investigate the structure of the cake layer, i.e., boundary layer of rejected particles, on membrane surfaces by Kim and Hoek (2002) and Chen et al. (2005), and a comprehensive review can be found elsewhere (Chen and Kim, 2006).

4. Concluding remarks and perspectives

With more than a half century history (Metropolis et al., 1953), Monte Carlo methods are widely used in science and engineering, focused on the properties and phase equilibria of polymers, colloids, proteins, biological and synthetic membranes, liquid crystals, semiconductors, solid-liquid and liquid-liquid interfaces, nano-materials, and materials for energy production and storage. Use of MC methods in environmental engineering is still in a burgeoning stage. However, because water and air are primary subjects of the discipline, MC simulations of particles in air and water or their interfaces can significantly contribute to a fundamental understanding of natural phenomena. All the MC simulations are, mathematically speaking, numerical integrations of large dimensions. Specific phenomenological modeling does not need to include rigor and formalism of statistical physics. Transition probability leads a system of random initial configuration to an equilibrium state. Efficient simulation requires advanced sampling techniques using biased probability and distributed parallel computing can be readily employed for rapid MC simulations.

In the future, it is expected that rapid and accurate MC algorithms will be generated and that these algorithms are readily applicable to real situations. Specific ensembles and thermodynamic properties will be updated as MC algorithms develop. More unified MC algorithms will be developed for easier use and a multi-scale version will be available. A more efficient sampling method is desired by using bias schemes which follows basic statistical physics as well as carefully mimicking phenomena of interest.

5. Appendix

5.1 Additional analysis of Monty Hall Dilemma

A complex way of calculating $P\left(S_1 \cap H_3\right)$ of Eq. (16) is as follows.

$$\begin{aligned} P\left(S_1 \cap H_3\right) &= P\left(S_1 \cap H_3 | C_2\right) + P\left(S_1 \cap H_3 | \widetilde{C_2}\right) && (100) \\ &= P\left(S_1 \cap H_3 | C_2\right) + P\left(S_1 \cap H_3 | C_1 \text{ or } C_3\right) && (101) \\ &= P\left(S_1 \cap H_3 | C_2\right) + P\left(S_1 \cap H_3 | C_1\right) + P\left(S_1 \cap H_3 | C_3\right) && (102) \\ &= P\left(S_1 \cap H_3 | C_2\right) + P\left(S_1 \cap H_3 | C_1\right) && (103) \end{aligned}$$

where $P\left(S_1 \cap H_3 | C_3\right) = 0$ because if the car is behind door 3, then the host will not open door 3. Further calculations provide

$$\begin{aligned} P\left(S_1 \cap H_3 | C_2\right) &= P\left(S_1 | C_2\right) P\left(H_3 | C_2\right) P\left(C_2\right) && (104) \\ &= P\left(S_1\right) P\left(H_3 | C_2\right) P\left(C_2\right) && (105) \\ &= \frac{1}{3} \cdot 1 \cdot \frac{1}{3} = \frac{1}{9} && (106) \end{aligned}$$

where $P\left(H_3 | C_2\right) = 1$ because the host has to open door #3 given that the car is behind door #2 after we selected door #1, and

$$\begin{aligned} P\left(S_1 \cap H_3 | C_1\right) &= P\left(S_1 | C_1\right) P\left(H_3 | C_1\right) P\left(C_1\right) \\ &= P\left(S_1\right) P\left(H_3 | C_1\right) P\left(C_1\right) \\ &= \frac{1}{3} \cdot \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{18} \end{aligned}$$

where $P\left(H_3 | C_1\right) = 1/2$ because the host can open either door #3 or #2 if I selected door #1. Therefore, as expected, we here obtain the identical answer to Eq. (16):

$$P\left(S_1 \cap H_2\right) = \frac{1}{9} + \frac{1}{18} = \frac{3}{18} = \frac{1}{6} \qquad (107)$$

6. References

M. P. Allen and D. J. Tildesley. *Computer Simulation of Liquids*. Oxford University Press, Oxford, UK, 1987.

G. Arya and A. Z. Panagiotopoulos. Monte carlo study of shear-induced alignment of cylindrical micelles in thin films. *Phys. Rev. E*, 70(3):031501, 2004.

M. Bayes and M. Price. An essay towards solving a problem in the doctrine of chances. *Phil. Trans. Roy. Soc. London*, 53:370–418, 1763.

G. Box and M. Muller. A note on the generation of random normal deviates. *The Annals of Mathematical Statistics*, 29(2):610–611, 1958.

R. Brown. A brief account of microscopical observations. *Phil. J.*, pages 358–371, 1828.

J. C. Chen and A. S. Kim. Brownian dynamics, molecular dynamics, and monte carlo modeling of colloidal systems. *Adv. Colloid Interface Sci.*, 112(1-3):159–173, 2004.

J. C. Chen and A. S. Kim. Monte carlo simulation of colloidal membrane filtration: Principal issues for modeling. *Adv. Colloid Interface Sci.*, 119(1):35–53, 2006.

J. C. Chen, M. Elimelech, and A. S. Kim. Monte carlo simulation of colloidal membrane filtration: Model development with application to characterization of colloid phase transition. *J. Membr. Sci.*, 255(1-2):291–305, 2005.

P. de Gennes. Reptation of a polymer chain in the presence of fixed obstacles. *J. Chem. Phys.*, 55(2):572–579, 1971.

J. de Pablo, M. Laso, and U. Suter. Simulation of polyethylene above and below the melting point. *J. Chem. Phys.*, 96(3):2395–2403, 1992.

D. Earl and M. Deem. Parallel tempering: Theory, applications, and new perspectives. *Phys. Chem. Chem. Phys.*, 7(23):3910–3916, 2005.

R. F. Feynman, R. B. Leighton, and M. Sands. *The Feynman Lectures on Physics*. Addison-Wesley Publishing Company, 1963.

A. Giffin. *Maximum Entropy: The Universal Method for Inference*. PhD thesis, University of Albany, Department of Physics, 2008.

A. Giffin. From physics to economics: An econometric example using maximum relative entropy. *Physica A: Statistical Mechanics and its Applications*, 388(8):1610–1620, 2009.

J. Happel. Viscous flow in multiparticle systems: slow motion of fluids relative to beds of spherical particles. *AIChE J.*, 4(2):197–201, 1958.

A. S. Kim and E. M. V. Hoek. Cake structure in dead-end membrane filtration: Monte carlo simulations. *Environ. Eng. Sci.*, 19(6):373–386, 2002.

A. S. Kim, S. Bhattacharjee, and M. Elimelech. Shear-induced reorganization of deformable molecular assemblages: Monte carlo studies. *Langmuir*, 17(2):552–561, 2001.

N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equation of state calculations by fast computing machines. *J. Chem. Phys.*, 21(6):1087–1092, 1953.

F. Mosteller. *Fifty Challenging Problems in Probability with Solutions*. Dover Publications, New York, 1965.

S. Northrup and J. McCammon. Simulation methods for protein structure fluctuations. *Biopolymers*, 19(5):1001–1016, 1980.

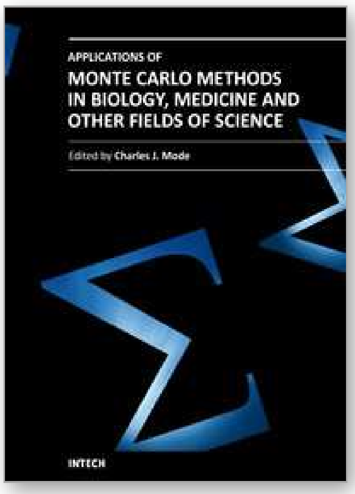
C. Pangali, M. Rao, and B. J. Berne. On a novel Monte Carlo scheme for simulating water and aqueous solutions. *Chem. Phys. Lett.*, 55(3):413–417, 1978.

F. Reif. *Fundamentals of Statistical and Thermal Physics*. McGraw-Hill, New York, 1965.

M. Rosenbluth and A. Rosenbluth. Monte carlo calculation of the average extension of molecular chains. *J. Chem. Phys.*, 23(2):356–359, 1955.

P. J. Rossky, J. Doll, and H. Friedman. Brownian dynamics as smart Monte Carlo simulation. *J. Chem. Phys.*, 69(10):4628–4633, 1978.

J. Siepmann and D. Frenkel. Configurational bias monte carlo: a new sampling scheme for flexible chains. *Mol. Phys.*, 75(1):59 – 70, 1992.



Applications of Monte Carlo Methods in Biology, Medicine and Other Fields of Science
Edited by Prof. Charles J. Mode

ISBN 978-953-307-427-6
Hard cover, 424 pages
Publisher InTech

This volume is an eclectic mix of applications of Monte Carlo methods in many fields of research should not be surprising, because of the ubiquitous use of these methods in many fields of human endeavor. In an attempt to focus attention on a manageable set of applications, the main thrust of this book is to emphasize applications of Monte Carlo simulation methods in biology and medicine.

Albert S. Kim (2011). Bias Monte Carlo Methods in Environmental Engineering, Applications of Monte Carlo Methods in Biology, Medicine and Other Fields of Science, Prof. Charles J. Mode (Ed.), ISBN: 978-953-307-427-6, InTech, Available from: http://www.intechopen.com/books/applications-of-monte-carlo-methods-in-biology-medicine-and-other-fields-of-science/bias-monte-carlo-methods-in-environmental-engineering

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821