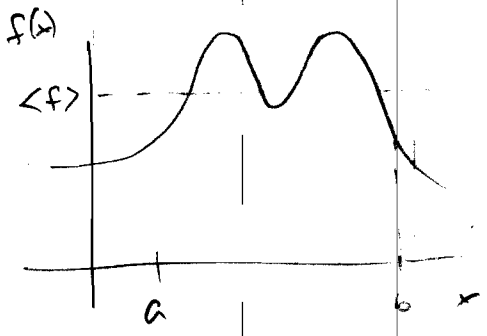
Monte Carlo Integrations and The Metropolis Algorithm

- Recall the concept of Importance Sampling

let say we have $\int_a^b f(x)\,dx \simeq (b-a)\langle f\rangle_x$ ← just uniformly sampled



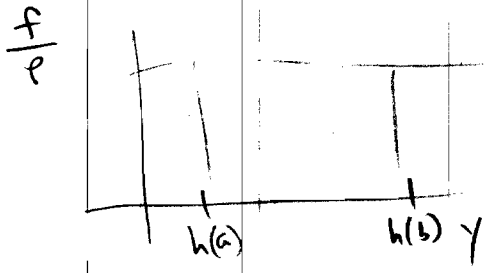
★ $(b-a)\langle f\rangle_{MC}$ might not be a good estimate of the true area under the curve due to the mis-sampling of $f(x)$ at humps & dips.

Now, if one can find a dist. function $\rho(x) \simeq f(x)$

so that we have $y = \int^x \rho(x')\,dx' = h(x)$

$$\int_a^b f(x)\,dx = \int_a^b \frac{f(x)}{\rho(x)}\,\rho(x)\,dx$$

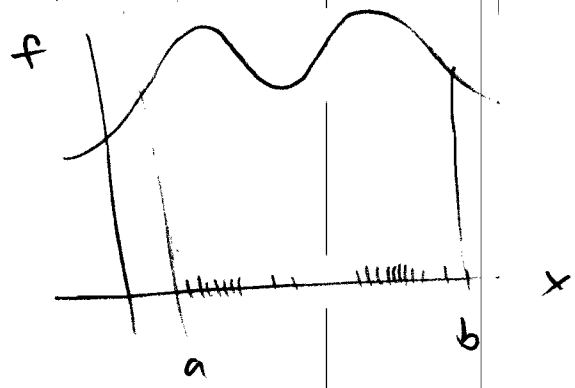
$$= \int_{h(a)}^{h(b)} \frac{f(h^{-1}(y))}{\rho(h^{-1}(y))}\,dy$$



In y-space, $(h(b) - h(a)) \left\langle \frac{f}{p} \right\rangle_y$ will

converge faster since $\frac{f}{p} \approx$ const in y-variable.

Another interpretation:



— the nonlinear coordinate transformation $h(x) = y = \int^x p(x')dx'$

is to concentrate more pts under the two humps

so as to adjust the sampling to put more

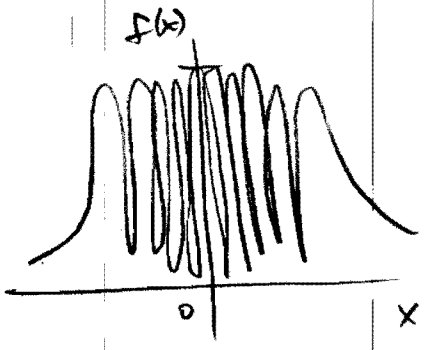
importance there!

Note: sampling uniformly in y ~ importance sampling in x

→ Similar to the transform method for generating nonuniform dist. from an uniform one. ☆ (For any arbitrary $f(x)$, transformation method does not work but we can use rejection method.)

— An example of using importance sampling to do an integral:

$I(x) = \int_a^b f(x')\,dx'$, where $f(x)$ is a badly behaving function



$f(x) = \sin^2 \frac{1}{x}$

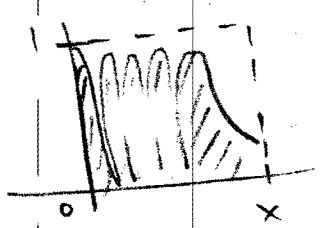
✡✡ SR, GQ won't work too well here!

✡ But, if we can find a way to use importance sampling to concentrate the x's under the humps, we can use MC to do this integral easily!

✡ Notice that since $0 < f(x) < 1$, we can use the rejection method to do this!

→ pick two independent uniform random #s (u, v) with $u \in [0,1]$ & $v \in [0, x]$

→ Since $0 < I(x) < x$ is the area under the curve given by $f(x)$,



the ratio of the areas $\frac{I(x)}{1 \cdot x}$ $\begin{matrix}(\text{under } f(x)) \\ (\text{total rect})\end{matrix}$ gives

the probability that a point (u,v) will be below the function $f(x)$!

→ And, this probability can be easily estimated by the rejected method :

$$\begin{cases} \text{keep } (u,v) \text{ if } & u \leq f(v) \\ \text{reject } (u,v) \text{ if } & u > f(v) \end{cases}$$

→ If M is the # of N points that were remained, then these M points will exactly be the points that are concentrated under $f(x)$ and

$$\text{prob} = \frac{M}{N} \simeq \frac{I(x)}{1 \cdot x}$$

this gives ⇒ $\underline{I(x) \simeq \frac{M}{N} x}$ for N large

(show graphs)

In order to achieve this we have to choose a dynamics for our simulation—a rule for changing from one state to another during the simulation—which results in each state appearing with exactly the probability appropriate to it. In the next chapter we will discuss at length a number of strategies for doing this, but the essential idea is that we try to simulate the physical processes that give rise to the master equation, Equation (1.1). We choose a set of rates $R(\mu \to \nu)$ for transitions from one state to another, and we choose them in such a way that the equilibrium solution to the corresponding master equation is precisely the Boltzmann distribution (1.5). Then we use these rates to choose the states which our simulated system passes through during the course of a simulation, and from these states we make estimates of whatever observable quantities we are interested in.

The advantage of this technique is that we need only sample quite a small fraction of the states of the system in order to get accurate estimates of physical quantities. For example, we do not need to include every state of the system in order to get a decent value for the partition function, as we would if we were to evaluate it directly from Equation (1.6). The principal disadvantage of the technique is that there are statistical errors in the calculation due to this same fact that we don't include every state in our calculation, but only some small fraction of the states. In particular this means that there will be statistical noise in the partition function. Taking the derivative of a noisy function is always problematic, so that calculating expectation values from derivatives of the partition function as discussed in Section 1.2 is usually not a good way to proceed. Instead it is normally better in Monte Carlo simulations to calculate as many expectations as we can directly, using equations such as (1.34). We can also make use of relations such as (1.36) to calculate quantities like susceptibilities without having to evaluate a derivative.

In the next chapter we will consider the theory of Monte Carlo simulation in equilibrium thermal systems, and the rest of the first part of the book will deal with the design of algorithms to investigate these systems. In the second part of the book we look at algorithms for non-equilibrium systems.

1.4 A brief history of the Monte Carlo method

In this section we outline the important historical developments in the evolution of the Monte Carlo method. This section is just for fun; feel free to skip over it to the next chapter if you're not interested.

The idea of Monte Carlo calculation is a lot older than the computer. The name "Monte Carlo" is relatively recent—it was coined by Nicolas Metropolis in 1949—but under the older name of "statistical sampling" the method has a history stretching back well into the last century, when numerical calculations were performed by hand using pencil and paper and perhaps

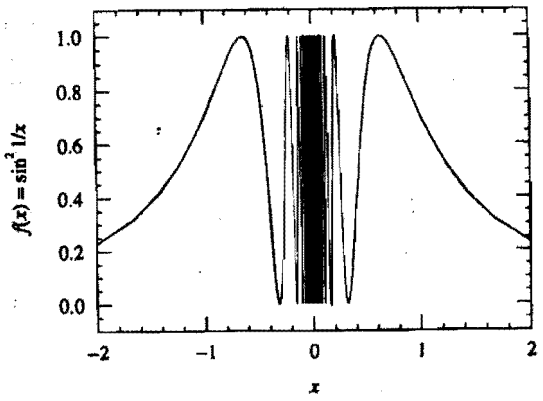


FIGURE 1.2 The pathological function $f(x) \equiv \sin^2 \frac{1}{x}$, whose integral with respect to x, though hard to evaluate analytically, can be evaluated in a straightforward manner using the Monte Carlo integration technique described in the text.

a slide-rule. As first envisaged, Monte Carlo was not a method for solving problems in physics, but a method for estimating integrals which could not be performed by other means. Integrals over poorly-behaved functions and integrals in high-dimensional spaces are two areas in which the method has traditionally proved profitable, and indeed it is still an important technique for problems of these types. To give an example, consider the function

$$f(x) \equiv \sin^2 \frac{1}{x} \tag{1.41}$$

which is pictured in Figure 1.2. The values of this function lie entirely between zero and one, but it is increasingly rapidly varying in the neighbourhood of $x = 0$. Clearly the integral

$$I(x) \equiv \int_0^x f(x')\,dx' \tag{1.42}$$

which is the area under this curve between 0 and x, takes a finite value somewhere in the range $0 < I(x) < x$, but it is not simple to calculate this value exactly because of the pathologies of the function near the origin. However, we can make an estimate of it by the following method. If we choose a random real number h, uniformly distributed between zero and x, and another v between zero and one and plot on Figure 1.2 the point for which these are the horizontal and vertical coordinates, the probability that

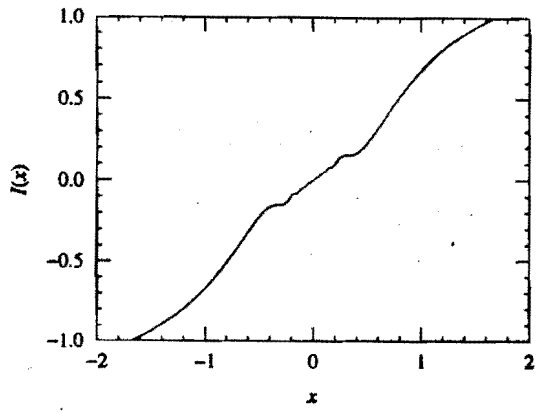


FIGURE 1.3 The function $I(x)$, calculated by Monte Carlo integration as described in the text.

this point will be below the line of $f(x)$ is just $I(x)/x$. It is easy to determine whether the point is in fact below the line: it is below it if $h < f(v)$. Thus if we simply pick a large number N of these random points and count up the number M which fall below the line, we can estimate $I(x)$ from

$$I(x) = \lim_{N\to\infty} \frac{Mx}{N}. \tag{1.43}$$

You can get an answer accurate to one figure by taking a thousand points, which would be about the limit of what one could have reasonably done in the days before computers. Nowadays, even a cheap desktop computer can comfortably run through a million points in a few seconds, giving an answer accurate to about three figures. In Figure 1.3 we have plotted the results of such a calculation for a range of values of x. The errors in this calculation are smaller than the width of the line in the figure.[7]

A famous early example of this type of calculation is the experiment known as "Buffon's needle" (Dörrie 1965), in which the mathematical constant π is determined by repeatedly dropping a needle onto a sheet of paper ruled with evenly spaced lines. The experiment is named after Georges-Louis Leclerc, Comte de Buffon who in 1777 was the first to show that if we throw a needle of length l completely at random onto a sheet of paper ruled with lines a distance d apart, then the chances that the needle will fall so as to

[7] In fact there exist a number of more sophisticated Monte Carlo integration techniques which give more accurate answers than the simple "hit or miss" method we have described here. A discussion can be found in the book by Kalos and Whitlock (1986).

intersect one of the lines is $2l/\pi d$, provided that $d \geq l$. It was Laplace in 1820 who then pointed out that if the needle is thrown down N times and is observed to land on a line M of those times, we can make an estimate of π from

$$\pi = \lim_{N\to\infty} \frac{2Nl}{Md}. \tag{1.44}$$

(Perhaps the connection between this and the Monte Carlo evaluation of integrals is not immediately apparent, but it will certainly become clear if you try to derive Equation (1.44) for yourself, or if you follow Dörrie's derivation.) A number of investigators made use of this method over the years to calculate approximate values for π. The most famous of these is Mario Lazzarini, who in 1901 announced that he had calculated a value of 3.1415929 for π from an experiment in which a $2\frac{1}{2}$ cm needle was dropped 3408 times onto a sheet of paper ruled with lines 3 cm apart. This value, accurate to better than three parts in ten million, would be an impressive example of the power of the statistical sampling method were it not for the fact that it is almost certainly faked. Badger (1994) has demonstrated extremely convincingly that, even supposing Lazzarini had the technology at his disposal to measure the length of his needle and the spaces between his lines to a few parts in 10^7 (a step necessary to ensure the accuracy of Equation (1.44)), still the chances of his finding the results he did were poorer than three in a million; Lazzarini was imprudent enough to publish details of the progress of the experiment through the 3408 castings of the needle, and it turns out that the statistical "fluctuations" in the numbers of intersections of the needle with the ruled lines are much smaller than one would expect in a real experiment. All indications are that Lazzarini forged his results. However, other, less well known attempts at the experiment were certainly genuine, and yielded reasonable figures for π: 3.1596 (Wolf 1850), 3.1553 (Smith 1855). Apparently, performing the Buffon's needle experiment was for a while quite a sophisticated pastime amongst Europe's intellectual gentry.

With the advent of mechanical calculating machines at the end of the nineteenth century, numerical methods took a large step forward. These machines increased enormously the number and reliability of the arithmetic operations that could be performed in a numerical "experiment", and made the application of statistical sampling techniques to research problems in physics a realistic possibility for the first time. An early example of what was effectively a Monte Carlo calculation of the motion and collision of the molecules in a gas was described by William Thomson (later Lord Kelvin) in 1901. Thomson's calculations were aimed at demonstrating the truth of the equipartition theorem for the internal energy of a classical system. However, after the fashion of the time, he did not perform the laborious analysis himself, and a lot of the credit for the results must go to Thomson's

of importance sampling

— The idea can be extended to any weighted average:

- here $g(x)$ can be any function.
- $g \triangleq \frac{f}{p}$ previously

$$\longrightarrow \int_a^b g(x)\, p(x)\, dx = \langle g \rangle_p$$ ← weighted average of g according to the dist. $p(x)$.

where $p(x)$ is a well-defined prob den with $\int_a^b p(x)\,dx = 1$.

In statistical physics & Quantum Mechanics,

we commonly encounter these averages:

— Ensemble averages of physical quantities such as E, S, T, ...

$$\langle E \rangle = \int E\, p(\underset{\sim}{x})\, d\underset{\sim}{x}$$

← $\underset{\sim}{x}$ is the configuration over phase space.

where $p(\underset{\sim}{x}) \sim e^{-E(\underset{\sim}{x})/kT}$ is the Boltzman's Distribution

— Quantum expectations

where $p(\underset{\sim}{x}) = \psi^*(x)\,\psi(x)$

← wave function of a particular system.

★ With an appropriate $p(x)$ determined

(in stat mech, it will be the Boltzman's dist.)

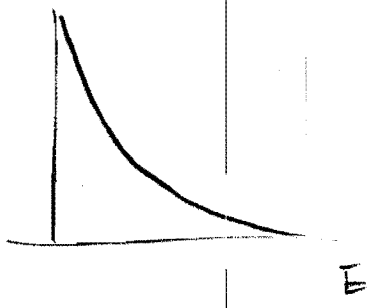
(otherwise, need some qualitative knowledge of $f(x)$),

the difficulty is in performing the weighted average

according to $p(x)$ <u>efficiently</u>, i.e. generating $\{x\}$ according to $p(x)$.

<u>Note</u>: The basic rejection method works but it might not be the most efficient.

e.g. $p(E) \sim e^{-E/kT}$ (standard Boltzman's)



For $E>0$, $p(E)$ is exponentially small with large E!

So, one will most likely generate configuration $\underline{x}$ such that the probability in rejecting it might be increasing large for G large.

→ A very powerful and efficient method to calculate these weighted averages is the Metropolis Algorithm.

Important Ingredients of Metropolis Algorithm:

① Replace spatial/ensemble average with (regardless) a temporal average of an equivalent (Markov) process that at equilibrium will produce the same invariant distribution of phase-space configurations $\rho(\underset{\sim}{x})$.

imitate nature

spatial/ensemble average → temporal average

randomly sample pts in phase space according to $\rho(\underset{\sim}{x})$ = sampling of phase space pts by a Markov process at equilibrium $P(\underset{\sim}{x} \to \underset{\sim}{x}')$

Note: for them to describe the same physical system, $\rho(\underset{\sim}{x})$ will be related to $P(\underset{\sim}{x} \to \underset{\sim}{x}')$.

Markov process : a prescription on how to generate the next conf. $\underset{\sim}{x}'$ from the previous conf. $\underset{\sim}{x}$ according to a transition probability

$$P(\underset{\sim}{x} \to \underset{\sim}{x}').$$

① $P(\underset{\sim}{x} \to \underset{\sim}{x}')$ is time invariant.

② $P(\underset{\sim}{x} \to \underset{\sim}{x}')$ depends on $\underset{\sim}{x}$ & $\underset{\sim}{x}'$ only.

Surely, we need

$$\int P(\underset{\sim}{x} \to \underset{\sim}{x}')\, d\underset{\sim}{x}' = 1.$$

(2) Detailed Balance

Recall that the process is assumed to be in equilibrium !

$\Rightarrow$ the rate at which the system make transitions into and out of state $\underset{\sim}{x}$ must be equal.

$$(*) \quad \sum_{\underset{\sim}{x}} P(\underset{\sim}{x})\, P(\underset{\sim}{x} \to \underset{\sim}{x}') = \sum_{\underset{\sim}{x}'} P(\underset{\sim}{x}')\, P(\underset{\sim}{x}' \to \underset{\sim}{x})$$

(out of $\underset{\sim}{x}$) (into $\underset{\sim}{x}$)

$\left\{ \begin{array}{l} P(\underset{\sim}{x}) \text{ - probability in state } \underset{\sim}{x} \\ P(\underset{\sim}{x} \Rightarrow \underset{\sim}{x}') \text{ - transition probability from } \underset{\sim}{x} \text{ to } \underset{\sim}{x}' \\ \sum_{x'} \text{ - summing of all states } \underset{\sim}{x}' \text{ that system can get from state } \underset{\sim}{x}. \end{array} \right.$

★ It turns out that an <u>invariant</u> $P(x)$ can't be uniquely generated by a Markov process using $P(x \to x')$ which satisfies (*) only!

[see Nerman P. 37 : it allows limit cycle solutions.]

★ What we can do is to enforce a <u>stronger</u> version of the equilibrium condition:

$$P(\underset{\sim}{x})\, P(\underset{\sim}{x} \to \underset{\sim}{x}') = P(\underset{\sim}{x}')\, P(\underset{\sim}{x}' \to \underset{\sim}{x})$$

(<u>detailed balance</u>)

★ microscopic dynamics has time invariant → one should expect detailed balance.

<u>Note</u> : No sum over $\underset{\sim}{x}'$!

— it is <u>more restrictive</u> in the sense that :

→ it is not the <u>sum</u> of all in & out rates cancels but the in & out rate for each <u>individual</u> process must cancel !

☆ One can show that if we have

① $P(\underline{x})\, P(\underline{x} \to \underline{x}') = P(\underline{x}')\, P(\underline{x}' \to \underline{x})$

and ② $\sum_{\underline{x}'} P(\underline{x} \to \underline{x}') = 1$

☆ then $P(\underline{x} \to \underline{x}')$ will uniquely determine the required invariant dist. $P(\underline{x})$.

Summary of our Goal: To define a sequence of Markov processes given by $P(\underline{x} \to \underline{x}')$ so that the resulting sequence of states $\{\underline{x}\}$ will be distributed according to $P(\underline{x})$.

→ The constraints on $P(\underline{x} \to \underline{x}')$ that will accomplish this goal are ① & ② !

Steps in a Metropolis Algorithm

e.g. N molecules in 3D

(1) Start with a system configuration $\underset{\sim}{X}$ ← $(\underset{\sim}{P}_{3N}, \underset{\sim}{Q}_{3N})$

(2) Construct a new system configuration $\underset{\sim}{X}'$ by a random process:

$$\underset{\sim}{X}' = \underset{\sim}{X} + \Delta\underset{\sim}{X}$$

(☆ Ising: just pick one random spin site)

where $\Delta\underset{\sim}{X}$ is a vector with random components from an uniform distribution.

(3) Compute the dist. function for $\underset{\sim}{X}$ & $\underset{\sim}{X}'$

$$P(\underset{\sim}{X}) \text{ & } P(\underset{\sim}{X}')$$

("prob" to be in $\underset{\sim}{X}$ & $\underset{\sim}{X}'$)

[Note: We know the functional form of $P(x)$ in our weighted average $\int P(\underset{\sim}{x})\, g(\underset{\sim}{x})\, d\underset{\sim}{x}$.
(The difficulty is to get $\{\underset{\sim}{x}\}$ that has $P(\underset{\sim}{x})$.)]

(4) To accept or reject the random choice $\underset{\sim}{X}'$ based on $P(\underset{\sim}{X} \to \underset{\sim}{X}')$ satisfying:

$$\begin{cases} P(\underset{\sim}{X})\, P(\underset{\sim}{X} \to \underset{\sim}{X}') = P(\underset{\sim}{X}')\, P(\underset{\sim}{X}' \to \underset{\sim}{X}) & (*) \\ \int P(\underset{\sim}{X} \to \underset{\sim}{X}')\, d\underset{\sim}{X} = 1 & (**) \end{cases}$$

(Note: only the ratio $\frac{P(\underset{\sim}{X} \to \underset{\sim}{X}')}{P(\underset{\sim}{X}' \to \underset{\sim}{X})} = \frac{P(\underset{\sim}{X}')}{P(\underset{\sim}{X})}$ is constrained by (*)!

We can write

$P(\underset{\sim}{x} \to \underset{\sim}{x}')$ as two terms:

$$P(\underset{\sim}{x} \to \underset{\sim}{x}') = \underbrace{w(\underset{\sim}{x}, \underset{\sim}{x}')}_{\text{selection Probability}} \underbrace{A(\underset{\sim}{x} \to \underset{\sim}{x}')}_{\text{acceptance probability}}$$

For simplicity, we assume that $w(\underset{\sim}{x}, \underset{\sim}{x}')$ is symmetric:

$$w(\underset{\sim}{x}, \underset{\sim}{x}') = w(\underset{\sim}{x}', \underset{\sim}{x})$$

Then, ⊛ (detailed balance) gives:

$$\frac{P(\underset{\sim}{x} \to \underset{\sim}{x}')}{P(\underset{\sim}{x}' \to \underset{\sim}{x})} = \frac{w(\underset{\sim}{x}, \underset{\sim}{x}')}{w(\underset{\sim}{x}', \underset{\sim}{x})} \frac{A(\underset{\sim}{x} \to \underset{\sim}{x}')}{A(\underset{\sim}{x}' \to \underset{\sim}{x})} = \frac{A(\underset{\sim}{x} \to \underset{\sim}{x}')}{A(\underset{\sim}{x}' \to \underset{\sim}{x})} = \frac{P(\underset{\sim}{x}')}{P(\underset{\sim}{x})}$$

So, the constraint on the ratio on P's

→ constraint on the ratio of A's.

✡✡ Note: We have freedom to choose $w(\underset{\sim}{x}, \underset{\sim}{x}')$ & the form of $A(\underset{\sim}{x} \to \underset{\sim}{x}')$ to satisfy both ⊛ & (✡✡)!

Main point

★★ Since we want our Markov process to be efficient in making changes to the configuration $\underset{\sim}{x}$,

→ we want to make $A(\underset{\sim}{x} \to \underset{\sim}{x}')$ as large as possible while (*) & (**) can be satisfied!

One way to make this choice for $A(\underset{\sim}{x} \to \underset{\sim}{x}')$ is the metropolis choice:

$$\begin{cases} \text{If } \rho(\underset{\sim}{x}') \geq \rho(\underset{\sim}{x}), & \text{take } A(\underset{\sim}{x} \to \underset{\sim}{x}') = 1 \\ \text{If } \rho(\underset{\sim}{x}') < \rho(\underset{\sim}{x}), & \text{take } A(\underset{\sim}{x} \to \underset{\sim}{x}') = \dfrac{\rho(\underset{\sim}{x}')}{\rho(\underset{\sim}{x})} < 1 \end{cases}$$

Note: To satisfy (*), in principle, $A(\underset{\sim}{x}' \to \underset{\sim}{x})$ can be adjusted!

reverse acceptance prob.

To satisfy (**), in principle, $w(\underset{\sim}{x}, \underset{\sim}{x}')$ can be adjusted!

→ We know that we can but we don't need to actually solve for them since we only need $A(\underset{\sim}{x} \to \underset{\sim}{x}')$.

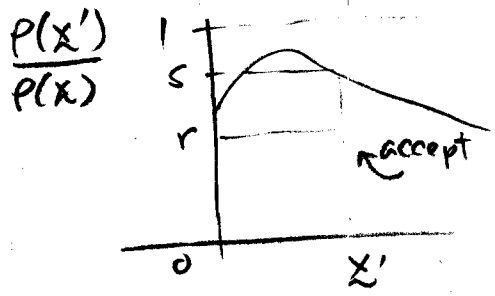
$A(\underline{x} \to \underline{x}') = 1$ means that we will always accept the new step.

$A(\underline{x} \to \underline{x}') = \frac{P(\underline{x}')}{P(\underline{x})} = s$ means that we will accept the new step with probability s!

To do this, we pick a random # $r \in [0,1]$ (like flipping a weighted coin) from an uniform distribution,

$\begin{cases} \text{if } r \le s, & \text{we accept the new step} \\ \text{if } r > s, & \text{we stay with the old step.} \end{cases}$

Note: this is similar to the rejection method



(3) Evaluate $g(\underset{\sim}{x})$ and

form sum $\sum g(\underset{\sim}{x})$

over all metropolis steps (moved & stayed)

Note: $\sum$ should be form only after transient has been taken out.

(6) repeat (2). ($\{\underset{\sim}{x}\}$ generated from this will dist according to $P(\underset{\sim}{x})$.)

(1) Evaluate the final weight average

$$\langle g \rangle_P = \int P(\underset{\sim}{x})\, g(\underset{\sim}{x})\, d\underset{\sim}{x} \underline{\underline{\text{by}}} \quad \frac{1}{M} \sum g(\underset{\sim}{x})$$

simple average over metro steps!

M - # of Metropolis steps taken.

Final reminder:

(1) If ⊛ (detailed balance) & (**) (normalization), then one can show that (we haven't don't it here)

$P(\underset{\sim}{x} \to \underset{\sim}{x}')$ uniquely determines $\rho(\underset{\sim}{x})$.

(random Markov transition process)

(2) Then, Metroplois algorithm is an efficient method to generate a sequence of Markov processes that will give $\{\underset{\sim}{x}\}$ with distribution $\rho(\underset{\sim}{x})$.

(3) This is basically a statistic process, the dist. of $\underset{\sim}{x}$ will approach $\rho(\underset{\sim}{x})$ as time $\to \infty$

$\Rightarrow$ we need to drop the transient.

(4) For almost all statistic physics problems, we have $\rho(\underset{\sim}{x}) \sim e^{-E(\underset{\sim}{x})/kT}$

(Boltzman's distribution)