Introduction to the stochastic evaluation of path integrals.

Why is this necessary?

- First of all, we should say that stochastic calculations (i.e. calculations involving sequences of random numbers) are a subset of numerical calculations, so we should first ask ourselves: Are numerical methods needed?

  The answer is a resounding 'YES!', but one should always keep in mind that in research numerical methods are not just to obtain numbers, but to provide insight where analytic methods **fail**. For us the prototypical examples will be strongly coupled many-body quantum systems, or, roughly speaking problems where neither perturbation theory nor saddle point approximations help.

- Ok, now why 'stochastic'? Here we follow Negele & Orland (Chapter 8, p. 400) Suppose we want to compute an equilibrium property $O(\{\vec{r}_i\})$ of a gas of $N = 200$ particles in a volume that we subdivide into 100 points per direction, ie. $V = \ell^3 100^3$, ℓ is the lattice spacing, then our quantity $O(\{\vec{r}_i\})$ will depend on $100^{3\times 200} = 10^{1200}$ variables. This number is so large that there is no computer in the universe that could store it or perform operations w/ it in any reasonable ($< 10^{10}$ y) amount of time. ☺

  It is for this reason that one introduces stochastic methods, in which the dimension of the integral becomes a secondary problem (Example below)

- Part of the issue here is also that we are attacking a formidable many-body problem with a ridiculous technique: trying to solve the Schrödinger equation for 200 particles, when all we may need is, e.g. the average density, sounds just like the well-known premise of statistical mechanics: we really don't care about all the "microscopic" details. This doesn't mean that we are neglecting something, there's no approximation involved, it just means that we don't need that much information.
In quantum mechanics, the right formalism for many-particle systems is of course many-body quantum theory, which can be formulated in terms of operators (second quantization) or fields (path integrals). (not op-valued)

We've been following the latter because we believe is the easiest one to work with. As it turns out, it is also the best path toward the numerical (in particular stochastic) solution of the quantum many-body problem.

- Before reformulating quantum mechanics for many particles, let's look at the problem of one particle, and see how stochastic methods become important in that case.

Consider the problem of finding the ground-state expectation value of a quantity $\hat{O}(\vec{x})$ (e.g. $\hat{x}^n$, $\hat{p}^n$) for a particle of mass m in a potential $V(\vec{x})$; i.e. we want to compute

$$\langle \hat{O}(\vec{x}) \rangle = \lim_{\beta \to \infty} \underbrace{\frac{\sum_n \langle n | \hat{O} e^{-\beta \hat{H}} | n \rangle}{\sum_n \langle n | e^{-\beta \hat{H}} | n \rangle}}_{\langle \hat{O}(x) \rangle_\beta} \qquad \hat{H} = \frac{\hat{p}^2}{2m} + V(\vec{x})$$

We know that we can write this as (before taking the $\beta \to \infty$ limit)

$$\langle \hat{O}(x) \rangle_\beta = \frac{1}{Z} \int_{x(0)=x(\beta)} D[\vec{x}(\tau)]\, e^{-\int_0^\beta d\tau \left[\frac{m}{2}\left(\frac{d\vec{x}}{d\tau}\right)^2 + V(\vec{x}(\tau))\right]} O(x)$$

where $$Z = \int_{x(0)=x(\beta)} D[\vec{x}(\tau)]\, e^{-\int_0^\beta d\tau \left[\frac{m}{2}\left(\frac{dx}{d\tau}\right)^2 + V(x(\tau))\right]}$$

To fix ideas you can imagine that $O(x) = x^2$, for example.

Now, how do we solve this problem? We know that the τ direction should actually be thought of as discrete points, so let's start there:

$$Z = \int_{-\infty}^{\infty} \prod_{k=1}^{N_\tau} d^3x_k \; e^{-\varepsilon \sum_{k=1}^{N_\tau} \left[\frac{m}{2}\frac{(\vec{x}_k - \vec{x}_{k-1})^2}{\varepsilon^2} + V(\vec{x}_k)\right]}$$

$N_\tau \cdot \varepsilon = \beta$

$d\tau \to \varepsilon$

$\int \to \sum$

... and we have a $3N_\tau$ dimensional integral, where eventually we want to take the limit $N_\tau \to \infty$.

- One more time: this is where stochastic methods come in.

- Quick questions for everyone:
  - Do you see anything peculiar if $V(\vec{x}) = \frac{1}{2} k \vec{x}^2$?
  - Can you solve the problem in that case?

The case of $V(\vec{x}) = \frac{k}{2}\vec{x}^2$ is of course the harmonic oscillator. What you need to do is write the exponent as $y^T M y$ by changing variables, and then you obtain a gaussian integral, which we all know how to do.

But what about other potentials, such as

$$V(\vec{x}) = \frac{a}{x^6} - \frac{b}{x^{12}}$$

or $V(\vec{x}) = \frac{k}{2}x^2 + \frac{\lambda}{4}x^4$, etc. (?)

One more time: where perturbation theory or saddle point fails, we need to look at this numerically, and we need stochastic methods for that.

One of the most important (or perhaps the most important?) theorems in the field of stochastic methods is the...

- Central limit theorem

Given a variable $\vec{x}$ distributed according to a probability $P(\vec{x})$ (assumed normalized $\int P(x)dx = 1$; $P(x) \geq 0$), the probability distribution of the mean $X = \sum_{i=1}^{N} f(\vec{x}_i)$ converges to a gaussian distribution when $N \to \infty$.

($\vec{x}_i \in P(\vec{x})$ (indep samples!))

(don't confuse N particles w/ N samples!)

$$P(X) \xrightarrow[N\to\infty]{} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(X-\bar{X})^2}{2\sigma^2}}$$

, $\sigma = \sqrt{\langle f^2 \rangle_P - \langle f \rangle_P^2} \times \frac{1}{\sqrt{N}}$

$\langle g \rangle_P = \int d^d x \, g(x) P(x)$

Proof of the central limit theorem.

Assume $P(x) \geq 0$

$$\int P(\vec{x})\, d^d x = 1$$

(When averaging over N indep. samples.)

The probability of obtaining a particular value X when $\vec{x}$ is distributed according to $P(x)$ may be written as:

$$\Pi(X) = \int \prod_{i=1}^{N} d^d x_i \, P(\vec{x}_i) \, \delta\left(\frac{1}{N} \sum_{j=1}^{N} f(\vec{x}_j) - X \right)$$

= "Sum over probabilities of simultaneously obtaining N values of x such that $X = \frac{1}{N}\sum_{j=1}^{N} f(x_j)$"

→ they are indep samples, so we just take the product.

Now we represent the δ function as a Fourier transform.

$$\Pi(X) = \int \prod_i d^d x_i \, P(x_i) \int \frac{N d\lambda}{2\pi} \, e^{iN\lambda X - i\lambda \sum_{j=1}^{N} f(x_j)} = \ldots = \frac{N}{2\pi} \int d\lambda \, e^{N F(\lambda, X)}$$

$$F(\lambda, X) = i\lambda X + g(\lambda)$$

$$g(\lambda) = \ln\left[\int d^d y \, e^{-i\lambda f(\vec{y})} P(\vec{y}) \right]$$

N is large !

Next: fix X, do saddle point evaluation of λ integral, using the stationarity condition $\left.\frac{\partial F}{\partial \lambda}\right|_{\tilde{\lambda}(X)} = 0$ $\quad \tilde{\lambda}$ = stationary point.

Notice that $\left.\frac{d^2 F}{d\lambda^2}\right|_{\tilde{\lambda}} = g''(\tilde{\lambda}(X))$

Thus, $$\Pi(X) = \left(\frac{N}{-g''(\tilde{\lambda}(X))\, 2\pi} \right)^{1/2} e^{N F(\tilde{\lambda}(X), X)} \left(1 + O(1/N) \right)$$

From the ~~stat~~ cond. we get

$$\left.\frac{\partial F}{\partial \lambda}\right|_{\tilde{\lambda}(X)} = iX + g'(\tilde{\lambda}) = 0 \quad \Rightarrow \quad X = i g'(\tilde{\lambda})$$

$$= \frac{\int d^N y\, f(y)\, e^{-i\tilde{\lambda} f}\, P}{\int d^N y\, e^{-i\tilde{\lambda} f}\, P}$$

which implicitly defines $\tilde{\lambda}(X)$.

To see how P depends on X more explicitly let us find the extrema:

$$0 = \frac{dF(\tilde{\lambda}, X)}{dX} = \frac{\partial F}{\partial X} + \frac{\partial F}{\partial \lambda}\frac{\partial \tilde{\lambda}}{\partial X} = i\tilde{\lambda}(X) \quad \left(\text{used } \left.\frac{\partial F}{\partial \lambda}\right|_{\tilde{\lambda}} = 0\right)$$

$$\Rightarrow X = \langle f \rangle_P$$

Moreover,

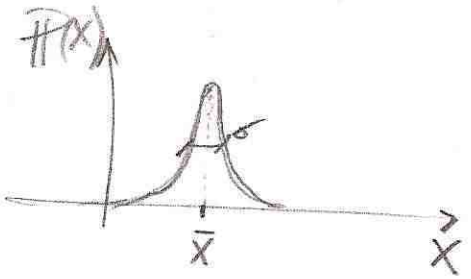
$$\frac{d^2 F}{dX^2} = \frac{1}{g''(\lambda=0)} = \frac{-1}{\langle f^2 \rangle_P - \langle f \rangle_P^2} < 0$$

$P(X)$ thus has a single maximum at $X = \langle f \rangle_P$ and monotonically decreases w/ curvature $\langle f^2 \rangle_P - \langle f \rangle_P^2$. As $N \to \infty$ the higher order terms contribute less and less to the tails of the distribution around X, and we approach

$$P(X) \xrightarrow[N\to\infty]{} \sqrt{\frac{N}{2\pi(\langle f^2 \rangle_P - \langle f \rangle_P^2)}}\; e^{\frac{-(X - \langle f \rangle_P)^2}{\frac{2}{N}\left(\langle f^2 \rangle_P - \langle f \rangle_P^2\right)}}$$

This is quite powerful if you think about it (it says "averaging makes sense")
It implies that you can approximate

$$\int d^d x \; f(\vec{x}) P(\vec{x}) \approx \frac{1}{N} \sum_{\substack{i=1 \\ x_i \in P}}^{N} f(x_i) \pm \sigma$$



arbitrary !!

where we can estimate $\sigma^2 \simeq \frac{1}{N}\left[\frac{1}{N}\sum_i f^2(x_i) - \left(\frac{1}{N}\sum_i f(x_i)\right)^2\right]$ (if $N \gg 1$)

- This by itself is what makes Monte Carlo methods work... and fail!
When $f(x)$ oscillates wildly, such that the signal $(\bar{X})$ to noise (σ) ratio is small, the method fails.
This happens, unfortunately, in many systems of interest (relativistic theories at finite μ, fermions w/ repulsive interactions, etc.)

At this point you can probably see where this is going:

- If I could (somehow) generate N independent configurations $\{\vec{x}_i\}$ distributed according to $P(\vec{x}) = \frac{\exp[-S_E]}{Z}$, $S_E = \varepsilon \sum \left(\frac{m}{2}\dot{x}^2 + V(x)\right)$
then I would be able to estimate

$$\langle \hat{O} \rangle_\beta \approx \sum_{i=1}^{N} O(x_i) \pm \frac{\sigma}{\sqrt{N}}$$ (regardless of dimension!)

The problem of MC simulations is fundamentally this one: how to efficiently generate configurations when $S_E[x]$ is extremely non-linear & non-local.

- Sampling strategies

Given $S_E[x]$, how can we obtain, as efficiently as possible, N independent (as far as possible) configurations of $\{x_i\}$?

- Remember:

I will tell you about 4 possible strategies:

1. Heat-bath algorithm
2. Metropolis algorithm
3. Molecular dynamics algorithm
4. Hybrid (2+3) algorithms (there are also 1+2 hybrids!)

1. The idea of the heat-bath algorithm is simple: if you can, use the exact probability to generate random configurations. Typically, generators are based on uniform random number generators (RNG). If your RNG is very good (i.e. successive numbers are independent according to correlation tests) then your samples will be independent, and you are in business!

Of course, this is only applicable in certain peculiar cases, e.g. studies of quenched QED, or other simple pure gauge theories (i.e. no fermions!).

(See e.g. Creutz, Phys. Rev. D 21, 2308 (1980)
Kennedy & Pendleton, Phys. Lett B 156, 393 (1985))

In our present simple case the bottom line is: if you can find a way to sample $S_E[x]$ exactly in a non-trivial case (non-trivial $V(x)$) use it!

Methods ②, ③ and ④ also sample the probability exactly, but they do it by resorting to a concept I have not discussed yet: Markov chains.

At the conceptual level, Markov chains are a recipe to generate random numbers according to a certain distribution, via a step-by-step procedure that takes you from one configuration to the next, starting from an arbitrary point.

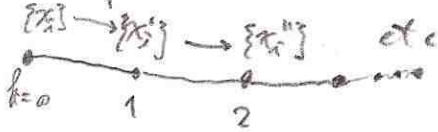
From a practical point of view there are two very important issues to take into account when using Markov chains:

ⓐ It takes a certain "thermalization time" for the chain to equilibrate and "lose memory" of the starting configuration, after which the samples start to obey the desired probability.

ⓑ Consecutive configurations in the chain are extremely correlated with each other, which means we have to wait for a certain "decorrelation time" when picking samples to do our stochastic estimation of the path integral.

Comments:

- Do not confuse Markov time with imaginary time!



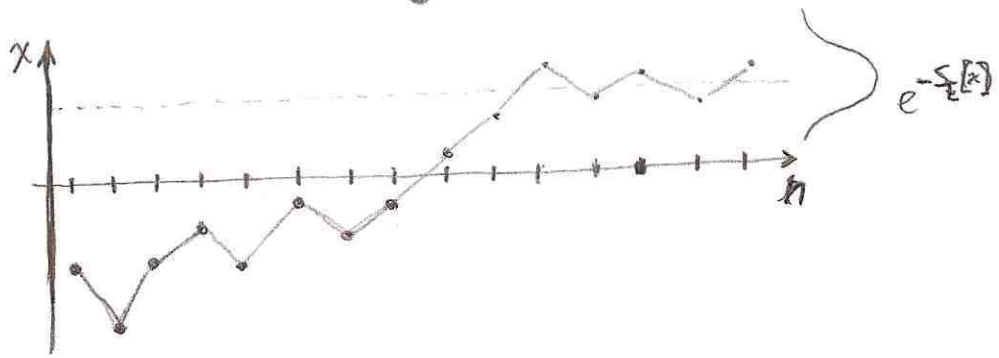
- Interestingly, both thermalization and decorrelation can be particularly slow close to phase transitions ("critical slowing down

A Markov chain is a "rule" which specifies the probability distribution for the $(n+1)^{th}$ element $x^{(n+1)}$ based only on the n^{th} element $x^{(n)}$

We will denote with $P(x \to y)$ the probability to obtain y starting from x.

We are interested in algorithms that generate, using Markov chains, a distribution of x's that converges to a specified distribution $P = e^{-S_E[x]}$.

Example



For a Markov process to sample $e^{-S[x]}$ it is sufficient that the rule $P(x \to y)$ has eventual access to every point in configuration space (ergodicity), and that it satisfies the reversibility condition:

$$e^{-S[x]} P(x \to y) = e^{-S[y]} P(y \to x)$$

Indeed, notice that e^{-S} is an equilibrium solution:

$$P(y) = \int dx\, e^{-S[x]} P(x \to y) = e^{-S[y]} \underbrace{\int dx\, P(y \to x)}_{= 1 \text{ (ergodicity!)}} = e^{-S[y]}$$

Is this equilibrium stable?

Suppose that the distribution at some step is $M(x)$.

The deviation from equilibrium is

$$D = \int dx \left| M(x) - e^{-S[x]} \right|$$

and at the next step it is

$$D' = \int dy \left| \underbrace{\int dx\, M(x) P(x \to y)}_{M'(y)} - e^{-S[y]} \right|$$

But then

$$D' = \int dy \left| \int dx \left(M(x) P(x \to y) - e^{-S[x]} P(x \to y) \right) \right| = \overbrace{\int dy\, P(x \to y)}^{1} \left| \int dx \left(M(x) - e^{-S[x]} \right) \right|$$

$$\leq \int dx \left| M(x) - e^{-S[x]} \right| = D$$

So $M(x) \longrightarrow e^{-S[x]}$. QED

Great! Now how do we construct a rule $P(x \to y)$?

Perhaps the most popular algorithm to do this is the Metropolis algorithm.

- The recipe is the following:

0. Pick a starting configuration x
1. Pick a new configuration x^T (T = tentative)
2. Compute $q = e^{-[S[x^T] - S[x]]}$
3. Pick a uniform random number $\xi \in [0, 1]$
4. If $\xi < q$, set $x = x^T$; If $\xi > q$, retain x.
   (Notice that if $S[x^T] < S[x]$, $q > 1 \Rightarrow$ lower action implies acceptance).
5. Repeat from 1.

If we place no restrictions on how we pick the candidate x_s, all we need to do is check reversibility. This is easy:

Going from x to y,

- If $e^{-S(y)} > e^{-S(x)}$, $x \to y$ is accepted; and
  $y \to x$ is accepted w/ probability $q = \dfrac{e^{-S(x)}}{e^{-S(y)}} = e^{-(S(x)-S(y))}$

Therefore, $\dfrac{P(x\to y)}{P(y\to x)} = \dfrac{1}{e^{-S(x)}/e^{-S(y)}} = \dfrac{e^{-S(y)}}{e^{-S(x)}}$ ✓

- Similarly one proves for $e^{-S(x)} \geq e^{-S(y)}$

We have assumed that the procedure to build x^T from x is symmetric, ie. there is no preferred direction $x \to x^T$ or $x^T \to x$ a priori.