

The Monte Carlo Method in Quantum Field Theory

Colin Morningstar

Carnegie Mellon University

HUGS June 2006



Outline

- Path integrals in quantum mechanics
- Monte Carlo integration and Markov chains
- Monte Carlo evaluation of path integral in quantum mechanics
- Free Klein-Gordon scalar field theory in $2 + 1$ dimensions
- Interacting ϕ^4 scalar field theory in $2 + 1$ dimensions
- Applications in quantum chromodynamics
- Topics for future study

Part I

Path integrals in quantum mechanics

Transition amplitudes in quantum mechanics

- key quantity in quantum mechanics: **transition amplitude**

$$Z(b, a) \equiv \langle x_b(t_b) | x_a(t_a) \rangle$$

- $Z(b, a)$ is probability amplitude for particle to go from point x_a at time t_a to point x_b at time t_b
- in this talk, will work in Heisenberg picture
 - state vectors $|\Psi\rangle$ are stationary
 - operators and their eigenvectors evolve with time

$$x(t) = e^{iHt/\hbar} x(0) e^{-iHt/\hbar}$$

$$|x(t)\rangle = e^{iHt/\hbar} |x(0)\rangle$$

- often will shift Hamiltonian so ground state energy is zero

$$H |\phi_n(t)\rangle = E_n |\phi_n(t)\rangle, \quad E_0 = 0$$

$$|\phi_0(t)\rangle = |\phi_0(0)\rangle \equiv |0\rangle$$

Spectral representation of transition amplitude

- insert complete (discrete) set of Heisenberg-picture eigenstates $|\phi_n(t)\rangle$ of Hamiltonian H into transition amplitude

$$Z(b, a) \equiv \langle x_b(t_b) | x_a(t_a) \rangle = \sum \langle x_b(t_b) | \phi_n(t_b) \rangle \langle \phi_n(t_b) | x_a(t_a) \rangle$$

- now use $|\phi_n(t)\rangle = e^{iHt/\hbar} |\phi_n(0)\rangle = e^{iE_n t/\hbar} |\phi_n(0)\rangle$ to obtain

$$Z(b, a) = \sum e^{-iE_n(t_b-t_a)/\hbar} \langle x_b(t_b) | \phi_n(t_b) \rangle \langle \phi_n(t_a) | x_a(t_a) \rangle$$

- finally, $\langle x(t) | \phi_n(t) \rangle \equiv \varphi_n(x)$ is the wavefunction in coordinate space, so

$$Z(b, a) = \sum_n \varphi_n(x_b) \varphi_n^*(x_a) e^{-iE_n(t_b-t_a)/\hbar}$$

- transition amplitude contains information about all energy levels and all wavefunctions \rightarrow spectral representation

Vacuum saturation

- take $t_a = -T$ and $t_b = T$ in the limit $T \rightarrow (1 - i\epsilon)\infty$

$$\begin{aligned}\langle x_b(T) | x_a(-T) \rangle &= \langle x_b(0) | e^{-iHT/\hbar} e^{iH(-T)/\hbar} | x_a(0) \rangle \\ &= \sum_{n=0}^{\infty} \langle x_b(0) | \phi_n(0) \rangle \langle \phi_n(0) | x_a(0) \rangle e^{-2iE_n T/\hbar} \\ &\rightarrow \langle x_b(0) | 0 \rangle \langle 0 | x_a(0) \rangle\end{aligned}$$

- insert complete set of energy eigenstates, use $E_{n+1} \geq E_n$, $E_0 = 0$, assume nondegenerate vacuum
- possibility of probing ground state (vacuum) properties

Vacuum expectation values

- now apply limit $T \rightarrow (1 - i\epsilon)\infty$ to more complicated amplitude

$$\begin{aligned} & \langle x_b(T) | x(t_2)x(t_1) | x_a(-T) \rangle \\ = & \langle x_b(0) | e^{-iHT/\hbar} x(t_2)x(t_1) e^{-iHT/\hbar} | x_a(0) \rangle \\ = & \sum_{n,m} \langle x_b(0) | \phi_n(0) \rangle \langle \phi_n(0) | x(t_2)x(t_1) | \phi_m(0) \rangle \langle \phi_m(0) | x_a(0) \rangle \\ & \times e^{-i(E_n+E_m)T/\hbar} \\ \rightarrow & \langle x_b(0) | 0 \rangle \langle 0 | x(t_2)x(t_1) | 0 \rangle \langle 0 | x_a(0) \rangle \end{aligned}$$

- hence, vacuum expectation values from

$$\langle 0 | x(t_2)x(t_1) | 0 \rangle = \lim_{T \rightarrow (1-i\epsilon)\infty} \frac{\langle x_b(T) | x(t_2)x(t_1) | x_a(-T) \rangle}{\langle x_b(T) | x_a(-T) \rangle}$$

- result generalizes to higher products of position operator

Observables from correlation functions

- all observables can be extracted from the correlation functions (vacuum expectation values)
- example: energies of the stationary states

$$\begin{aligned}\langle 0|x(t)x(0)|0\rangle &= \langle 0|e^{iHt/\hbar}x(0)e^{-iHt/\hbar}x(0)|0\rangle \\ &= \sum_n \langle 0|x(0)e^{-iHt/\hbar}|\phi_n(0)\rangle \langle \phi_n(0)|x(0)|0\rangle \\ &= \sum_n |\langle 0|x(0)|\phi_n(0)\rangle|^2 e^{-iE_n t/\hbar}\end{aligned}$$

- similarly for more complicated correlation functions

$$\begin{aligned}\langle 0|x^2(t)x^2(0)|0\rangle &= \langle 0|e^{iHt/\hbar}x^2(0)e^{-iHt/\hbar}x^2(0)|0\rangle \\ &= \sum_n |\langle 0|x^2(0)|\phi_n(0)\rangle|^2 e^{-iE_n t/\hbar}\end{aligned}$$

- but difficult to extract energies E_n from above oscillatory functions
→ much easier if we had **decaying** exponentials

The imaginary time formalism

- can get decaying exponentials if we rotate from the **real** to the **imaginary** axis in time (Wick rotation) $t \rightarrow -i\tau$

$$\langle 0|x(t)x(0)|0\rangle = \sum_n |\langle 0|x(0)|\phi_n(0)\rangle|^2 e^{-E_n\tau/\hbar}$$
$$\xrightarrow{\tau \rightarrow \infty} |\langle 0|x(0)|0\rangle|^2 + |\langle 0|x(0)|\phi_1(0)\rangle|^2 e^{-E_1\tau/\hbar}$$

- later, will see this imaginary time formalism provides another important advantage for Monte Carlo applications

Quantum mechanics and path integrals

- in the 1940s, Feynman developed an alternative formulation of quantum mechanics (his Ph.D. thesis)
 - Richard Feynman, *Rev Mod Phys* **20**, 367 (1948)

- quantum mechanical law of motion:
 - probability amplitude from *sum over histories*

$$Z(b, a) \sim \sum_{\substack{\text{all paths } x(t) \\ \text{from } a \text{ to } b}} \exp(iS[x(t)]/\hbar)$$



- all paths contribute to probability amplitude, but with different *phases* determined by the *action* $S[x(t)]$
- classical limit: when small changes in path yield changes in action large compared to \hbar , phases cancel out and path of least action $\delta S = 0$ dominates sum over histories

Defining the path integral

- action = time integral of Lagrangian (kinetic minus potential energy)

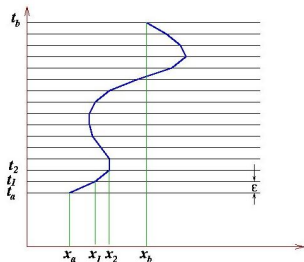
$$S = \int dt L(x, \dot{x}) = \int dt (K - U)$$

- divide time into steps of width ε where $N\varepsilon = t_b - t_a$
- path integral is defined as

$$Z(b, a) = \lim_{\varepsilon \rightarrow 0} \frac{1}{A} \int_{-\infty}^{\infty} \frac{dx_1}{A} \frac{dx_2}{A} \dots \frac{dx_{N-1}}{A} e^{iS[x(t)]/\hbar}$$

where A is a normalization factor depending on ε chosen so path integral well-defined

- in nonrelativistic theory, paths cannot double-back in time



Schrödinger equation

- probability amplitude $\psi(x_b, t_b)$ at time t_b given amplitude $\psi(x_a, t_a)$ at earlier time t_a given by

$$\psi(x_b, t_b) = \int Z(b, a) \psi(x_a, t_a) dx_a$$

- take $t_a = t$ and $t_b = t + \varepsilon$ one time slice away

$$\psi(x_b, t + \varepsilon) = \frac{1}{A} \int_{-\infty}^{\infty} \exp \left[\frac{i\varepsilon}{\hbar} L \left(\frac{x_b + x_a}{2}, \frac{x_b - x_a}{\varepsilon} \right) \right] \psi(x_a, t) dx_a$$

- in L , take $\dot{x} = (x_b - x_a)/\varepsilon$ and mid-point prescription $x \rightarrow (x_b + x_a)/2$
- particle in potential: $L = \frac{1}{2}m\dot{x}^2 - V(x, t)$, write $x_b = x$, $x_a = x + \eta$

$$\psi(x, t + \varepsilon) = \frac{1}{A} \int_{-\infty}^{\infty} e^{im\eta^2/(2\hbar\varepsilon)} e^{-i\varepsilon V(x+\eta/2, t)/\hbar} \psi(x + \eta, t) d\eta$$

Schrödinger equation (continued)

$$\psi(x, t + \varepsilon) = \frac{1}{A} \int_{-\infty}^{\infty} e^{im\eta^2/(2\hbar\varepsilon)} e^{-i\varepsilon V(x+\eta/2, t)/\hbar} \psi(x + \eta, t) d\eta$$

- rapid oscillation of $e^{im\eta^2/(2\hbar\varepsilon)}$ except when $\eta \sim O(\sqrt{\varepsilon}) \rightarrow$ integral dominated by contributions from η having values of this order
- expand to $O(\varepsilon)$ and $O(\eta^2)$, except $e^{im\eta^2/(2\hbar\varepsilon)}$ (ψ refers to $\psi(x, t)$)

$$\begin{aligned} \psi + \varepsilon \frac{\partial \psi}{\partial t} &= \frac{1}{A} \int_{-\infty}^{\infty} e^{im\eta^2/(2\hbar\varepsilon)} \left[1 - \frac{i\varepsilon}{\hbar} V(x, t) \right] \left[\psi + \eta \frac{\partial \psi}{\partial x} + \frac{\eta^2}{2} \frac{\partial^2 \psi}{\partial x^2} \right] d\eta \\ &= \frac{1}{A} \int_{-\infty}^{\infty} e^{im\eta^2/(2\hbar\varepsilon)} \left[\psi - \frac{i\varepsilon}{\hbar} V(x, t) \psi + \eta \frac{\partial \psi}{\partial x} + \frac{\eta^2}{2} \frac{\partial^2 \psi}{\partial x^2} \right] d\eta \end{aligned}$$

Schrödinger equation (continued)

$$\psi + \varepsilon \frac{\partial \psi}{\partial t} = \frac{1}{A} \int_{-\infty}^{\infty} e^{im\eta^2/(2\hbar\varepsilon)} \left[\psi - \frac{i\varepsilon}{\hbar} V(x,t)\psi + \eta \frac{\partial \psi}{\partial x} + \frac{\eta^2}{2} \frac{\partial^2 \psi}{\partial x^2} \right] d\eta$$

- matching leading terms on both sides determines A (analytic continuation to evaluate integral)

$$1 = \frac{1}{A} \int_{-\infty}^{\infty} e^{im\eta^2/(2\hbar\varepsilon)} d\eta = \frac{1}{A} \left(\frac{2\pi i\hbar\varepsilon}{m} \right)^{1/2} \Rightarrow A = \left(\frac{2\pi i\hbar\varepsilon}{m} \right)^{1/2}$$

- two more integrals:

$$\frac{1}{A} \int_{-\infty}^{\infty} e^{im\eta^2/(2\hbar\varepsilon)} \eta d\eta = 0, \quad \frac{1}{A} \int_{-\infty}^{\infty} e^{im\eta^2/(2\hbar\varepsilon)} \eta^2 d\eta = \frac{i\hbar\varepsilon}{m}$$

- $O(\varepsilon)$ part of equation at top yields

$$-\frac{\hbar}{i} \frac{\partial \psi}{\partial t} = -\frac{\hbar^2}{2m} \frac{\partial^2 \psi}{\partial x^2} + V(x,t)\psi$$

- the Schrödinger equation!

Free particle in one dimension

- Lagrangian of free particle in one dimension $L = \frac{1}{2}m\dot{x}^2$
- amplitude for particle to travel from x_a at time t_a to location x_b at later time t_b is

$$\langle x_b(t_b) | x_a(t_a) \rangle = \int_a^b \mathcal{D}x(t) \exp(iS[b, a]/\hbar)$$

summing over all allowed paths with $x(t_a) = x_a$ and $x(t_b) = x_b$.

- classical path $x_{cl}(t)$ from $\delta S = 0$ and boundary conditions:

$$\ddot{x}_{cl}(t) = 0, \quad x_{cl}(t) = x_a + (x_b - x_a) \frac{(t - t_a)}{(t_b - t_a)}$$

- classical action is

$$S_{cl}[b, a] = \int_{t_a}^{t_b} dt \frac{1}{2}m\dot{x}_{cl}^2 = \frac{m(x_b - x_a)^2}{2(t_b - t_a)}$$

- write $x(t) = x_{cl}(t) + \chi(t)$ where $\chi(t_a) = \chi(t_b) = 0$ then

$$S[b, a] = S_{cl}[b, a] + \int_{t_a}^{t_b} dt \frac{1}{2}m\dot{\chi}^2$$

where $S_{cl}[b, a]$ is classical action; no terms linear in $\chi(t)$ since S_{cl} is extremum

Path integral for free particle

- amplitude becomes ($T = t_b - t_a$)

$$Z(b, a) = F(T) \exp(iS_{cl}/\hbar)$$

$$F(T) = \int_0^T \mathcal{D}\chi \exp \left\{ \frac{im}{2\hbar} \int_0^T dt \dot{\chi}^2 \right\}$$

- partition time into discrete steps of length ε , use midpoint prescription, and note that $\chi_0 = \chi_N = 0$

$$\int_0^T \mathcal{D}\chi = \frac{1}{A} \int_{-\infty}^{\infty} \left(\prod_{l=1}^{N-1} \frac{d\chi_l}{A} \right) \quad A = \left(\frac{2\pi i \hbar \varepsilon}{m} \right)^{1/2}$$

$$\int_0^T dt \dot{\chi}^2 = \frac{1}{\varepsilon} \sum_{j=0}^{N-1} (\chi_{j+1} - \chi_j)^2$$

$$F(T) = \left(\frac{m}{2\pi i \hbar \varepsilon} \right)^{N/2} \int_{-\infty}^{\infty} \left(\prod_{l=1}^{N-1} d\chi_l \right) \exp \left\{ \frac{im}{2\hbar \varepsilon} \chi_j M_{jk} \chi_k \right\}$$

Gaussian integration

- a multivariate Gaussian integral remains

$$F(T) = \left(\frac{m}{2\pi i \hbar \varepsilon}\right)^{N/2} \int_{-\infty}^{\infty} \left(\prod_{l=1}^{N-1} d\chi_l\right) \exp\left\{\frac{im}{2\hbar\varepsilon} \chi_j M_{jk} \chi_k\right\}$$

where M is a symmetric $(N-1) \times (N-1)$ matrix

$$M = \begin{bmatrix} 2 & -1 & 0 & 0 & \cdots \\ -1 & 2 & -1 & 0 & \cdots \\ 0 & -1 & 2 & -1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

- Gaussian integrals of symmetric matrix A easily evaluated

$$\int_{-\infty}^{\infty} \left(\prod_{i=1}^n d\chi_i\right) \exp\left(-\chi_j A_{jk} \chi_k\right) = \left(\frac{\pi^n}{\det A}\right)^{1/2}$$

- result:

$$F(T) = \left(\frac{m}{2\pi i \hbar \varepsilon \det M}\right)^{1/2}$$

Determinant evaluation

- now need to compute $\det(M)$
- consider $n \times n$ matrix B_n of form

$$B_n = \begin{pmatrix} 2b & -b & 0 & 0 & \cdots \\ -b & 2b & -b & 0 & \cdots \\ 0 & -b & 2b & -b & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}_{n,n}$$

- notice that

$$\begin{aligned} \det B_n &= 2b \det B_{n-1} + b \det \left(\begin{array}{c|ccc} -b & -b & 0 & \cdots \\ 0 & & & \\ \vdots & & & \end{array} \right) \\ &= 2b \det B_{n-1} - b^2 \det B_{n-2} \end{aligned}$$

- define $I_n = \det B_n$ then have recursion relation

$$I_{n+1} = 2bI_n - b^2I_{n-1}, \quad I_{-1} = 0, \quad I_0 = 1, \quad n = 0, 1, 2, \dots$$

Transition amplitude for free particle

- rewrite $I_{n+1} = 2bI_n - b^2I_{n-1}$, $I_{-1} = 0$, $I_0 = 1$ as

$$\begin{pmatrix} I_{n+1} \\ I_n \end{pmatrix} = \begin{pmatrix} 2b & -b^2 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} I_n \\ I_{n-1} \end{pmatrix} = \begin{pmatrix} 2b & -b^2 \\ 1 & 0 \end{pmatrix}^n \begin{pmatrix} I_1 \\ I_0 \end{pmatrix}$$

- straightforward to show that

$$\begin{pmatrix} 2b & -b^2 \\ 1 & 0 \end{pmatrix}^n = \begin{pmatrix} (n+1)b^n & -nb^{n+1} \\ nb^{n-1} & -(n-1)b^n \end{pmatrix}$$

- so that

$$\begin{pmatrix} I_{n+1} \\ I_n \end{pmatrix} = \begin{pmatrix} (n+1)b^n & -nb^{n+1} \\ nb^{n-1} & -(n-1)b^n \end{pmatrix} \begin{pmatrix} 2b \\ 1 \end{pmatrix}$$

- and thus, $I_n = \det B_n = (n+1)b^n$

- here, $b = 1$ and $n = N - 1$ so $\det M = N$ and using $N\varepsilon = t_b - t_a$ obtain

$$F(t_b, t_a) = \left(\frac{m}{2\pi i \hbar (t_b - t_a)} \right)^{1/2}$$

- Final result:

$$\langle x_b(t_b) | x_a(t_a) \rangle = \left(\frac{m}{2\pi i \hbar (t_b - t_a)} \right)^{1/2} \exp \left\{ \frac{im(x_b - x_a)^2}{2\hbar(t_b - t_a)} \right\}$$

Infinite square well

- one of the first systems usually studied when learning quantum mechanics is the infinite square well
- particle moving in one dimension under influence of potential

$$V(x) = \begin{cases} 0 & \text{for } 0 < x < L \\ \infty & \text{for } x \leq 0 \text{ and } x \geq L \end{cases}$$

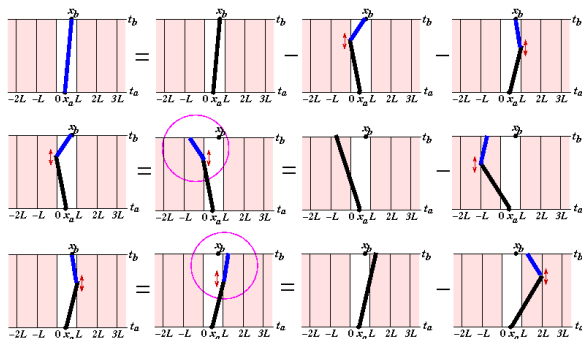
- path integral for transition amplitude given by

$$Z(b, a) = \lim_{\epsilon \rightarrow 0} \frac{1}{A} \int_0^L \frac{dx_1}{A} \cdots \int_0^L \frac{dx_{N-1}}{A} \exp \left\{ \frac{im}{2\epsilon\hbar} \sum_{j=0}^{N-1} (x_{j+1} - x_j)^2 \right\}$$

- paths limited to $0 < x < L$
- gaussian integrals over bounded domains produce error functions \rightarrow direct evaluation difficult in closed form
- extend regions of integration to $-\infty < x < \infty$, but subtract off forbidden paths
 - M. Goodman, Am. Jour. Phys. **49**, 9 (1981)

Path cancellations

- **black** lines: all **unbounded** paths between end points
- **blue** lines: paths between end points that do **not cross** an nL boundary
- no doubling back in time
- **magenta** circle indicates action preserving reflection

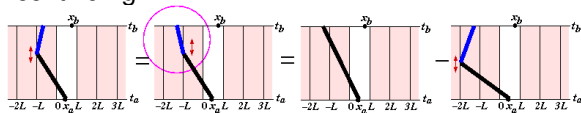


end point
 $-x_b$

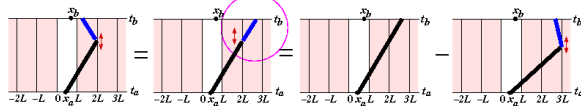
end point
 $2L - x_b$

Path cancellations (continued)

- continuing



end point
 $-2L + x_b$



end point
 $2L + x_b$

- and so on forever \rightarrow final result is

$$\begin{aligned}
 \langle x_b, t_b | x_a, t_a \rangle_{\text{well}} &= \langle x_b, t_b | x_a, t_a \rangle_{\text{free}} \\
 &\quad - \langle -x_b, t_b | x_a, t_a \rangle_{\text{free}} - \langle 2L - x_b, t_b | x_a, t_a \rangle_{\text{free}} \\
 &\quad + \langle -2L + x_b, t_b | x_a, t_a \rangle_{\text{free}} + \langle 2L + x_b, t_b | x_a, t_a \rangle_{\text{free}} + \dots \\
 &= \sum_{n=-\infty}^{\infty} \left\{ \langle 2nL + x_b, t_b | x_a, t_a \rangle_{\text{free}} - \langle 2nL - x_b, t_b | x_a, t_a \rangle_{\text{free}} \right\}
 \end{aligned}$$

Transition amplitude for infinite square well

- substitute amplitude for free particle

$$\langle x_b(t_b) | x_a(t_a) \rangle = \left(\frac{m}{2\pi i \hbar (t_b - t_a)} \right)^{1/2} \\ \times \sum_{n=-\infty}^{\infty} \left(\exp \left\{ \frac{im(2nL + x_b - x_a)^2}{2\hbar(t_b - t_a)} \right\} - \exp \left\{ \frac{im(2nL - x_b - x_a)^2}{2\hbar(t_b - t_a)} \right\} \right)$$

- apply Poisson summation and integrate the gaussian

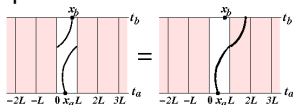
$$\sum_{n=-\infty}^{\infty} f(n) = \sum_{j=-\infty}^{\infty} \int_{-\infty}^{\infty} ds f(s) e^{2\pi i j s} \\ \int_{-\infty}^{\infty} ds \exp(-i\alpha s^2 \pm i\beta s) = \sqrt{\frac{\pi}{i\alpha}} \exp\left(\frac{i\beta^2}{4\alpha}\right)$$

- spectral representation of transition amplitude

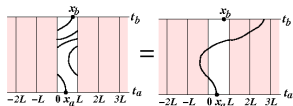
$$\langle x_b(t_b) | x_a(t_a) \rangle = \sum_{n=1}^{\infty} \varphi_n(x_b) \varphi_n^*(x_a) e^{-iE_n(t_b - t_a)/\hbar} \\ E_n = \frac{n^2 \pi^2 \hbar^2}{2mL^2} \quad \varphi_n(x) = \sqrt{\frac{2}{L}} \sin\left(\frac{n\pi x}{L}\right)$$

Free particle in 1D periodic box

- consider particle moving in one-dimension with periodic boundary conditions at $x = 0$ and $x = L$
- enforcing boundary conditions on paths difficult
- use trick similar to that used in infinite square well
- express set of allowed paths in terms of equivalent set of unrestricted paths



end point $x_b + L$



end point $x_b + 2L$

- result:

$$\langle x_b, t_b | x_a, t_a \rangle_{\text{periodic}} = \sum_{n=-\infty}^{\infty} \langle x_b + nL, t_b | x_a, t_a \rangle_{\text{free}}$$

Transition amplitude for periodic boundary

- substitute amplitude for free particle

$$\langle x_b(t_b) | x_a(t_a) \rangle = \left(\frac{m}{2\pi i \hbar (t_b - t_a)} \right)^{1/2} \sum_{n=-\infty}^{\infty} \exp \left\{ \frac{im(nL + x_b - x_a)^2}{2\hbar(t_b - t_a)} \right\}$$

- apply Poisson summation and integrate the gaussian

$$\sum_{n=-\infty}^{\infty} f(n) = \sum_{j=-\infty}^{\infty} \int_{-\infty}^{\infty} ds f(s) e^{2\pi i j s}$$
$$\int_{-\infty}^{\infty} ds \exp(-i\alpha s^2 \pm i\beta s) = \sqrt{\frac{\pi}{i\alpha}} \exp\left(\frac{i\beta^2}{4\alpha}\right)$$

- spectral representation of transition amplitude

$$\langle x_b(t_b) | x_a(t_a) \rangle = \sum_{n=-\infty}^{\infty} \varphi_n(x_b) \varphi_n^*(x_a) e^{-iE_n(t_b - t_a)/\hbar}$$

$$E_n = \frac{p_n^2}{2m} \quad p_n = \frac{2\pi n \hbar}{L} \quad \varphi_n(x) = \frac{1}{\sqrt{L}} e^{ip_n x / \hbar}$$

- quantization of momenta

The simple harmonic oscillator

- kinetic and potential energy of a simple harmonic oscillator of mass m and frequency ω

$$K = \frac{1}{2}m\dot{x}^2 \quad U = \frac{1}{2}m\omega^2x^2$$

- action is given by

$$S[x(t)] = \int_{t_a}^{t_b} dt \left(\frac{1}{2}m\dot{x}^2 - \frac{1}{2}m\omega^2x^2 \right)$$

- classical equations of motion

$$\delta S = 0 \quad \Rightarrow \quad \ddot{x}_{\text{cl}} + \omega^2x_{\text{cl}} = 0$$

- value of action for the classical path ($T = t_b - t_a$)

$$S_{\text{cl}} = \frac{m\omega}{2 \sin(\omega T)} \left[(x_a^2 + x_b^2) \cos(\omega T) - 2x_ax_b \right]$$

- to calculate, write path as deviation from classical path

$$x(t) = x_{\text{cl}}(t) + \chi(t) \quad \chi(t_a) = \chi(t_b) = 0$$

Path integral of simple harmonic oscillator

- amplitude can then be written as

$$Z(b, a) = F(T) \exp(iS_{cl}/\hbar)$$

$$F(T) = \int_0^0 \mathcal{D}\chi \exp \left\{ \frac{im}{2\hbar} \int_0^T dt (\dot{\chi}^2 - \omega^2 \chi^2) \right\}$$

- partition time into discrete steps of length ε and use midpoint prescription

$$\int_0^0 \mathcal{D}\chi = \frac{1}{A} \int_{-\infty}^{\infty} \left(\prod_{l=1}^{N-1} \frac{d\chi_l}{A} \right) \quad A = \left(\frac{2\pi i \hbar \varepsilon}{m} \right)^{1/2}$$

$$\int_0^T dt (\dot{\chi}^2 - \omega^2 \chi^2) = \frac{1}{\varepsilon} \sum_{j=0}^{N-1} \left[(\chi_{j+1} - \chi_j)^2 - \frac{\varepsilon^2 \omega^2}{4} (\chi_{j+1} + \chi_j)^2 \right]$$

$$F(T) = \left(\frac{m}{2\pi i \hbar \varepsilon} \right)^{N/2} \int_{-\infty}^{\infty} \left(\prod_{l=1}^{N-1} d\chi_l \right) \exp \left\{ \frac{im}{2\hbar \varepsilon} \chi_j \mathbf{M}_{jk} \chi_k \right\}$$

Gaussian integration

- a multivariate Gaussian integral remains

$$F(T) = \left(\frac{m}{2\pi i \hbar \varepsilon} \right)^{N/2} \int_{-\infty}^{\infty} \left(\prod_{l=1}^{N-1} d\chi_l \right) \exp \left\{ \frac{im}{2\hbar \varepsilon} \chi_j M_{jk} \chi_k \right\}$$

where M is a symmetric $(N-1) \times (N-1)$ matrix

$$M = \begin{bmatrix} 2 & -1 & 0 & 0 & \cdots \\ -1 & 2 & -1 & 0 & \cdots \\ 0 & -1 & 2 & -1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} - \frac{\varepsilon^2 \omega^2}{4} \begin{bmatrix} 2 & 1 & 0 & 0 & \cdots \\ 1 & 2 & 1 & 0 & \cdots \\ 0 & 1 & 2 & 1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

- Gaussian integrals are easily evaluated

$$F(T) = \left(\frac{m}{2\pi i \hbar \varepsilon \det M} \right)^{1/2}$$

Evaluating the determinant

- now must compute $\det M$
- consider $\det(B_n)$ where $n \times n$ matrix B_n has form

$$B_n = \begin{pmatrix} a & b & 0 & 0 & \cdots \\ b & a & b & 0 & \cdots \\ 0 & b & a & b & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}_{n,n}$$

- matches M for $n = N - 1$, $a = 2(1 - \epsilon^2\omega^2/4)$, $b = -(1 + \epsilon^2\omega^2/4)$
- notice that

$$\begin{aligned} \det B_n &= a \det B_{n-1} - b \det \left(\begin{array}{c|ccc} b & b & 0 & \cdots \\ \hline 0 & & & \\ \vdots & & & \\ & & B_{n-2} & \end{array} \right) \\ &= a \det B_{n-1} - b^2 \det B_{n-2} \end{aligned}$$

- define $I_n = \det B_n$ to obtain recursion relation

$$I_{n+1} = aI_n - b^2I_{n-1}, \quad I_{-1} = 0, \quad I_0 = 1, \quad n = 0, 1, 2, \dots$$

Evaluating the determinant (continued)

- rewrite recursion relation as

$$\begin{pmatrix} I_{n+1} \\ I_n \end{pmatrix} = \begin{pmatrix} a & -b^2 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} I_n \\ I_{n-1} \end{pmatrix} = \begin{pmatrix} a & -b^2 \\ 1 & 0 \end{pmatrix}^n \begin{pmatrix} I_1 \\ I_0 \end{pmatrix}$$

- diagonalize

$$\begin{pmatrix} a & -b^2 \\ 1 & 0 \end{pmatrix} = \mathcal{S} \begin{pmatrix} \lambda_+ & 0 \\ 0 & \lambda_- \end{pmatrix} \mathcal{S}^{-1}$$

$$\lambda_{\pm} = \frac{1}{2} \left(a \pm \sqrt{a^2 - 4b^2} \right),$$

$$\mathcal{S} = \begin{pmatrix} \lambda_+ & \lambda_- \\ 1 & 1 \end{pmatrix} \quad \mathcal{S}^{-1} = \frac{1}{\lambda_+ - \lambda_-} \begin{pmatrix} 1 & -\lambda_- \\ -1 & \lambda_+ \end{pmatrix}$$

- then we have

$$\begin{pmatrix} I_{n+1} \\ I_n \end{pmatrix} = \mathcal{S} \begin{pmatrix} \lambda_+^n & 0 \\ 0 & \lambda_-^n \end{pmatrix} \mathcal{S}^{-1} \begin{pmatrix} a \\ 1 \end{pmatrix}$$

- thus

$$I_n = \det B_n = \frac{\lambda_+^{n+1} - \lambda_-^{n+1}}{\lambda_+ - \lambda_-} \quad (\lambda_+ \neq \lambda_-)$$

Amplitude for simple harmonic oscillator

- using $\lambda_{\pm} = 1 \pm i\omega\epsilon + O(\epsilon^2)$ yields

$$\begin{aligned}\lim_{\substack{\epsilon \rightarrow 0 \\ N \rightarrow \infty}} \epsilon \det M &= \lim_{\substack{\epsilon \rightarrow 0 \\ N \rightarrow \infty}} \epsilon \frac{1}{2i\omega\epsilon} \left((1 + i\omega\epsilon)^N - (1 - i\omega\epsilon)^N \right) \\ &= \lim_{\substack{\epsilon \rightarrow 0 \\ N \rightarrow \infty}} \frac{1}{2i\omega} \left(\left(1 + \frac{i\omega T}{N} \right)^N - \left(1 - \frac{i\omega T}{N} \right)^N \right) \\ &= \frac{1}{2i\omega} (e^{i\omega T} - e^{-i\omega T}) = \frac{\sin \omega T}{\omega}.\end{aligned}$$

- final result for the path integral

$$\langle x_b(t_b) | x_a(t_a) \rangle_{\text{sho}} = \left(\frac{m\omega}{2\pi i \hbar \sin(\omega(t_b - t_a))} \right)^{1/2} \exp\{iS_{\text{cl}}/\hbar\}$$

Evolution of gaussian wave packet

- for initial wave packet at time $t_a = 0$ with probability dist.

$$|\phi(x_a, t_a)|^2 = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x_a - \bar{x})^2}{2\sigma^2}\right)$$

- probability amplitude at later time

$$\begin{aligned}\phi(x_b, t_b) &= \int_{-\infty}^{\infty} dx_a Z(b, a) \phi(x_a, 0) \\ &= \left(\frac{-im\omega(2\pi)^{-3/2}}{\hbar\sigma \sin(\omega t_b)}\right)^{1/2} \int_{-\infty}^{\infty} dx_a e^{iS_{cl}/\hbar} e^{-(x_a - \bar{x})^2/(4\sigma^2)}\end{aligned}$$

- final result for probability distribution: Gaussian with width s

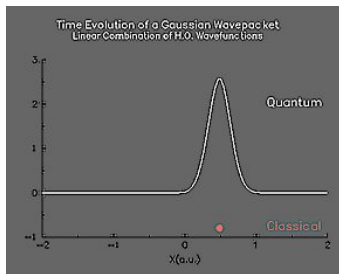
$$|\phi(x_b, t_b)|^2 = \frac{1}{s\sqrt{2\pi}} \exp\left(-\frac{(x_b - \bar{x} \cos(\omega t_b))^2}{2s^2}\right)$$

- new width given by

$$s = \sigma \left\{ \cos^2(\omega t_b) + \frac{\hbar^2}{4m^2\omega^2\sigma^4} \sin^2(\omega t_b) \right\}^{1/2}$$

Visualization

- time evolution of a Gaussian wave packet for a simple harmonic oscillator



mass $m = 1\text{g/mol} = 1.66 \times 10^{-27}\text{kg}$
frequency $\omega = 3 \times 10^{14}\text{radians/sec}$
initial wave packet:
center at 0.5 au
RMS spread 0.14 au
1 au (atomic unit) = 0.529 angstrom
probability distribution shown
(in inverse a.u.)

- completely calculated using path integrals → did **not** use Schrodinger equation

Other probability amplitudes

- so path integrals give us simple transition amplitudes

$$\langle x_b(t_b) | x_a(t_a) \rangle = \int_a^b \mathcal{D}x \exp \left\{ \frac{i}{\hbar} \int_{t_a}^{t_b} dt L(x, \dot{x}) \right\}$$

- but this important result generalizes to more complicated amplitudes

$$\begin{aligned} & \langle x_b(t_b) | x(t_2) x(t_1) | x_a(t_a) \rangle \\ &= \int_a^b \mathcal{D}x x(t_2)x(t_1) \exp \left\{ \frac{i}{\hbar} \int_{t_a}^{t_b} dt L(x, \dot{x}) \right\} \end{aligned}$$

for $t_a < t_1 < t_2 < t_b$

Path integrals in imaginary time

- in imaginary time formalism, paths contribute to sum over histories with real exponential weights (not phases)

$$\begin{aligned} & \langle x_b(\tau_b) | x(\tau_2) x(\tau_1) | x_a(\tau_a) \rangle \\ &= \int_a^b \mathcal{D}x x(\tau_2)x(\tau_1) \exp \left\{ -\frac{1}{\hbar} \int_{\tau_a}^{\tau_b} d\tau L(x, \dot{x}) \right\} \end{aligned}$$

- classical path gets highest weighting
- note that weights are all **real** and **positive** since action is real
 - this fact will be crucial for the Monte Carlo method

Vacuum expectation values from path integrals

- obtain correlation functions (vacuum expectation values) from ratios of path integrals

$$\begin{aligned}\langle 0|x(t_2)x(t_1)|0\rangle &= \lim_{T\rightarrow\infty} \frac{\langle x_b(T)|x(t_2)x(t_1)|x_a(-T)\rangle}{\langle x_b(T)|x_a(-T)\rangle} \\ &= \frac{\int_a^b \mathcal{D}x \, x(t_2)x(t_1) \exp\left\{-\frac{1}{\hbar} \int_{-\infty}^{\infty} d\tau L(x, \dot{x})\right\}}{\int_a^b \mathcal{D}x \exp\left\{-\frac{1}{\hbar} \int_{-\infty}^{\infty} d\tau L(x, \dot{x})\right\}}\end{aligned}$$

- generalizes to more complicated correlation functions
 - any correlation function can be computed using path integrals

Examples for the simple harmonic oscillator

- evaluating path integrals as before, the following correlation functions can be obtained ($\tau_1 \leq \tau_2 \leq \tau_3 \leq \tau_4$)

$$\begin{aligned}\langle 0|x(\tau_1)|0\rangle &= 0 \\ \langle 0|x(\tau_2)x(\tau_1)|0\rangle &= \frac{\hbar}{2m\omega} e^{-\omega(\tau_2-\tau_1)} \\ \langle 0|x(\tau_4)x(\tau_3)x(\tau_2)x(\tau_1)|0\rangle &= \left(\frac{\hbar}{2m\omega}\right)^2 e^{-\omega(\tau_4-\tau_1)} \\ &\quad \times \left[e^{-\omega(\tau_2-\tau_3)} + 2e^{-\omega(\tau_3-\tau_2)} \right]\end{aligned}$$

- comparison with spectral representation tells us

$$\begin{aligned}\langle 0|x(\tau)x(0)|0\rangle &= \frac{\hbar}{2m\omega} e^{-\omega\tau} \\ \Rightarrow E_1 - E_0 = \hbar\omega \quad |\langle 1|x(0)|0\rangle|^2 &= \frac{\hbar}{2m\omega}\end{aligned}$$

Another example in SHO

- excite vacuum with $x(\tau)^2$ operator

$$\langle 0|x^2(\tau)x^2(0)|0\rangle = \left(\frac{\hbar}{2m\omega}\right)^2 \left(1 + 2e^{-2\omega\tau}\right)$$

- compare with spectral representation at large time separations

$$\begin{aligned}\lim_{\tau \rightarrow \infty} \langle 0|x^2(\tau)x^2(0)|0\rangle &= |\langle 0|x^2(0)|0\rangle|^2 \\ &+ |\langle 2|x^2(0)|0\rangle|^2 e^{-(E_2-E_0)t/\hbar} + \dots \\ &= \left(\frac{\hbar}{2m\omega}\right)^2 \left(1 + 2e^{-2\omega\tau}\right)\end{aligned}$$

- interpretation:

$$E_2 - E_0 = 2\hbar\omega$$
$$|\langle 0|x^2(0)|0\rangle|^2 = \left(\frac{\hbar}{2m\omega}\right)^2 \quad |\langle 2|x^2(0)|0\rangle|^2 = 2 \left(\frac{\hbar}{2m\omega}\right)^2$$

One last example in SHO

- to determine expectation value of $x(0)^2$ in first-excited state

$$\langle 0|x(\tau) x^2(\frac{1}{2}\tau) x(0)|0\rangle = 3 \left(\frac{\hbar}{2m\omega} \right)^2 e^{-\omega\tau}$$

- compare with spectral interpretation at large times

$$\begin{aligned} \lim_{\tau \rightarrow \infty} \langle 0|x(\tau) x^2(\frac{1}{2}\tau) x(0)|0\rangle \\ = |\langle 0|x(0)|1\rangle|^2 \langle 1|x^2(0)|1\rangle e^{-(E_1-E_0)\tau/\hbar} + \dots \end{aligned}$$

- since $\langle 0|x(0)|0\rangle = \langle 0|x(\tau)|0\rangle = 0$
- by inspection and using previously derived results

$$\langle 1|x^2(0)|1\rangle = \frac{3\hbar}{2m\omega}$$

Pause for reflection

- observables in quantum mechanics can be extracted from the correlation functions (vacuum expectation values)
- imaginary time formalism is a great trick for assisting in such extractions
- correlation functions can be computed via path integrals

$$\begin{aligned} & \langle 0 | x(t_2) x(t_1) | 0 \rangle \\ = & \frac{\int_a^b \mathcal{D}x \, x(t_2) x(t_1) \exp \left\{ -\frac{1}{\hbar} \int_{-\infty}^{\infty} d\tau L(x, \dot{x}) \right\}}{\int_a^b \mathcal{D}x \, \exp \left\{ -\frac{1}{\hbar} \int_{-\infty}^{\infty} d\tau L(x, \dot{x}) \right\}} \end{aligned}$$



Part II

Monte Carlo integration and Markov chains

The die is cast?

- in rare situations, the path integrals can be computed exactly
 - simple harmonic oscillator, free particle
- sometimes the action can be written $S = S_0 + gS_I$
 - S_0 describes the free motion of the particles
 - path integrals using S_0 are Gaussian and can be exactly computed
 - S_I describes the interaction of the particles, but the coupling g is small
 - compute in perturbation theory as expansion in g
- however, if interactions are **not weak**
 - usually must resort to Monte Carlo methods
 - for example, quantum chromodynamics (QCD)

Simple Monte Carlo integration

- trapezoidal/Simpson's rule not feasible for integrals of very large dimension: too many function evaluations
- must start **gambling!**
- basic theorem of Monte Carlo integration

$$\int_V f(\vec{x}) d^D x \approx V \langle f \rangle \pm V \sqrt{\frac{\langle f^2 \rangle - \langle f \rangle^2}{N}}$$

$$\langle f \rangle \equiv \frac{1}{N} \sum_{i=1}^N f(\vec{x}_i) \qquad \langle f^2 \rangle \equiv \frac{1}{N} \sum_{i=1}^N f(\vec{x}_i)^2$$

- N points $\vec{x}_1, \dots, \vec{x}_N$ chosen **independently** and **randomly** with uniform probability dist. throughout D -dimensional volume V
- justified by the law of large numbers/central limit theorem
- in the limit $N \rightarrow \infty$, MC estimate tends to normal distribution, uncertainty tends to standard deviation

Quick review of probabilities

- consider an experiment whose outcome depends on chance
- represent an outcome by X called a **random variable**
- **sample space** Ω of experiment is set of all possible outcomes
- X is **discrete** if Ω is finite or countably infinite, **continuous** otherwise
- probability distribution for discrete X is real-valued function p_X on domain Ω satisfying $p_X(x) \geq 0$ for all $x \in \Omega$ and $\sum_{x \in \Omega} p_X(x) = 1$
- for any subset E of Ω , **probability** of E is $P(E) = \sum_{x \in E} p_X(x)$
- a sequence of random variables X_1, X_2, \dots, X_N that are mutually independent and have same distribution is called an **independent trials process**

Probability (continued)

- for continuous real-valued X , real-valued function p_X is a probability **density** and probability of outcome between real values a and b is $P(a \leq X \leq b) = \int_a^b p_X(s) ds$
- **cumulative** distribution is $F_X(x) = P(X \leq x) = \int_{-\infty}^x p_X(s) ds$
- common density: **normal** $p_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/(2\sigma^2)}$

Review: expected values

- **expected value** of X is

$$E(X) = \sum_{x \in \Omega} x p_X(x) \quad \left(= \int_{-\infty}^{\infty} s p_X(s) ds \right)$$

- properties: $E(X + Y) = E(X) + E(Y)$ and $E(cX) = cE(X)$
- for independent random variables X, Y have $E(XY) = E(X)E(Y)$
- can show $E(X)$ is average of outcomes if repeated many times
- for continuous real-valued function f , can show that

$$E(f(X)) = \sum_{x \in \Omega} f(x) p_X(x) \quad \left(= \int_{-\infty}^{\infty} f(s) p_X(s) ds \right)$$

- Proof: group together terms in $\sum_x f(x)p_X(x)$ having same $f(x)$ value
- denote set of different $f(x)$ values by \mathcal{F} , and subset of Ω leading to same value of $f(x)$ by $\Omega_{f(x)}$, then

$$\begin{aligned} \sum_{x \in \Omega} f(x) p_X(x) &= \sum_{y \in \mathcal{F}} \sum_{x \in \Omega_{f(x)}} f(x) p_X(x) = \sum_{y \in \mathcal{F}} y \left(\sum_{x \in \Omega_{f(x)}} p_X(x) \right) \\ &= \sum_{y \in \mathcal{F}} y p(y) = E(f(x)) \end{aligned}$$

Review: variances

- **variance** of X is $V(X) = E((X - E(X))^2)$
- **standard deviation** of X is $\sigma(X) = \sqrt{V(X)}$
- properties: $V(cX) = c^2V(X)$ and $V(X + c) = V(X)$
- for independent random variables X, Y have
 $V(X + Y) = V(X) + V(Y)$ (exercise: prove this)
- let X_1, \dots, X_N be an independent trials process with $E(X_j) = \mu$ and $V(X_j) = \sigma^2$, and define $A_N = (X_1 + X_2 + \dots + X_N)/N$, then can show
 $E(A_N) = \mu, \quad V(A_N) = \sigma^2/N$

Chebyshev inequality

- **Chebyshev inequality:** Let X be a discrete random variable with $E(X) = \mu$ and let $\epsilon > 0$ be any positive real number, then

$$P(|X - \mu| \geq \epsilon) \leq \frac{V(X)}{\epsilon^2}$$

- **Proof:**

- Let $p_X(x)$ denote distribution of X , then probability that X differs from μ by at least ϵ is $P(|X - \mu| \geq \epsilon) = \sum_{|x-\mu| \geq \epsilon} p_X(x)$

- considering positive summands and the ranges of summation,

$$V(X) = \sum_x (x - \mu)^2 p_X(x) \geq \sum_{|x-\mu| \geq \epsilon} (x - \mu)^2 p_X(x) \geq \sum_{|x-\mu| \geq \epsilon} \epsilon^2 p_X(x)$$

- but rightmost expression is

$$\epsilon^2 \sum_{|x-\mu| \geq \epsilon} p_X(x) = \epsilon^2 P(|X - \mu| \geq \epsilon)$$

- thus, have shown $V(x) \geq \epsilon^2 P(|X - \mu| \geq \epsilon)$

Weak law of large numbers

- **Weak law of large numbers:** Let X_1, X_2, \dots, X_N be an independent trials process with $E(X_j) = \mu$ and $V(X_j) = \sigma^2$, where μ, σ are finite, and let $A_N = (X_1 + X_2 + \dots + X_N)/N$. Then for any $\epsilon > 0$,

$$\lim_{N \rightarrow \infty} P(|A_N - \mu| \geq \epsilon) = 0, \quad \lim_{N \rightarrow \infty} P(|A_N - \mu| < \epsilon) = 1$$

- Proof:

- stated two slides ago that $E(A_N) = \mu$ and $V(A_N) = \sigma^2/N$
- from Chebyshev inequality

$$P(|A_N - \mu| \geq \epsilon) \leq \frac{V(A_N)}{\epsilon^2} = \frac{\sigma^2}{N\epsilon^2} \xrightarrow{N \rightarrow \infty} 0$$

- also known as the **law of averages**
- also applies to continuous random variables

Strong law of large numbers

- **Strong law of large numbers:** Let X_1, X_2, \dots, X_N be an independent trials process with $E(X_j) = \mu$ and $E(X_j^4) = K$, where μ, K are finite, then
$$P\left(\lim_{N \rightarrow \infty} (X_1 + X_2 + \dots + X_N)/N = \mu\right) = 1$$

- the finiteness of $E(X_j^4)$ is not needed, but simplifies proof

- Proof:

- define $Y_j = X_j - \mu$ so $E(Y_j) = 0$ and set $E(Y_j^4) = C < \infty$

- define $A_N = (Y_1 + Y_2 + \dots + Y_N)/N$

- given $E(Y_j) = 0$ and all Y_j are independent,

$$N^4 E(A_N^4) = NE(Y_j^4) + 6\binom{n}{2}E(Y_i^2 Y_j^2) = NC + 3N(N-1)E(Y_i^2)^2$$

- since $0 \leq V(Y_j^2) = E(Y_j^4) - E(Y_j^2)^2$ then $E(Y_j^2)^2 \leq E(Y_j^4) = C$

- so $E(A_N^4) \leq C/N^3 + 3C/N^2$ which means

$$E(\sum_{N=1}^{\infty} A_N^4) = \sum_{N=1}^{\infty} E(A_N^4) \leq \sum_{N=1}^{\infty} \left(\frac{C}{N^3} + \frac{3C}{N^2}\right) < \infty$$

- this implies $\sum_{N=1}^{\infty} A_N^4 < \infty$ with unit probability, and convergence of the series implies $\lim_{N \rightarrow \infty} A_N^4 = 0 \Rightarrow \lim_{N \rightarrow \infty} A_N = 0$

- proves $E(X)$ is average of outcomes for many repetitions

Application to one-dimensional integral

- if X is a random variable with probability density $p_X(x)$ and f is a well-behaved real-valued function, then $Y = f(X)$ is a random variable
- consider uniform density $p_X(x) = \begin{cases} 1/(b-a) & a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$
- use this probability density to obtain N outcomes X_1, X_2, \dots, X_n
- apply function f to obtain random variables $Y_j = f(X_j)$
- law of large numbers tell us that

$$\frac{1}{N} \sum_{j=1}^N Y_j \xrightarrow{N \rightarrow \infty} E(Y) = E(f(X)) = \frac{1}{(b-a)} \int_a^b f(s) ds$$

- define $\langle f \rangle \equiv \frac{1}{N} \sum_{j=1}^N f(X_j)$ then $(b-a) \lim_{N \rightarrow \infty} \langle f \rangle = \int_a^b f(s) ds$
- straightforward generalization to multiple dimensions
- how good is estimate for finite N ?

Central limit theorem

- **Central limit theorem:** Let X_1, X_2, \dots, X_N be independent random variables with common distribution having $E(X_j) = \mu$ and $V(X_j) = \sigma^2$, where μ, σ are finite, and let $A_N = (X_1 + X_2 + \dots + X_N)/N$. Then for $a < b$,

$$\lim_{N \rightarrow \infty} P \left(\frac{a\sigma}{\sqrt{N}} < (A_N - \mu) < \frac{b\sigma}{\sqrt{N}} \right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx$$

- alternatively: the distribution of $(X_1 + \dots + X_N - N\mu)/(\sigma\sqrt{N})$ tends to the standard normal (zero mean, unit variance)
- for proof, consult the literature
- for large N , the central limit theorem tells us that the error one makes in approximating $E(X)$ by A_N is $\sigma/\sqrt{N} = \sqrt{V(X)/N}$
- for $Y = f(X)$ as before, the error in approximating $E(f(X))$ by $\sum_j f(X_j)/N$ is $\sqrt{V(f(X))/N}$
- use Monte Carlo method to estimate $V(f(X))$

$$V(Y) = E((Y - E(Y))^2) \approx \langle (f - \langle f \rangle)^2 \rangle = \langle f^2 \rangle - \langle f \rangle^2$$

Application with non-uniform sampling

- if X is a random variable with probability density $p_X(x)$ and f is a well-behaved real-valued function, then $Y = f(X)$ is a random variable
- if $p_X(x)$ can be easily sampled, then use $p_X(x)$ to obtain N outcomes X_1, X_2, \dots, X_n
- apply function f to obtain random variables $Y_j = f(X_j)$
- law of large numbers tell us that

$$\frac{1}{N} \sum_{j=1}^N Y_j \xrightarrow{N \rightarrow \infty} E(Y) = E(f(X)) = \int_a^b p_X(s) f(s) ds$$

Monte Carlo integration

- recap of Monte Carlo integration (uniform sampling):

$$\int_V f(\vec{x}) d^D x \approx V \langle f \rangle \pm V \sqrt{\frac{\langle f^2 \rangle - \langle f \rangle^2}{N}}$$

$$\langle f \rangle \equiv \frac{1}{N} \sum_{i=1}^N f(\vec{x}_i) \qquad \langle f^2 \rangle \equiv \frac{1}{N} \sum_{i=1}^N f(\vec{x}_i)^2$$

- N points $\vec{x}_1, \dots, \vec{x}_N$ chosen **independently** and **randomly** with uniform probability dist. throughout D -dimensional volume V
- law of large numbers justifies correctness of estimate
- central limit theorem gives estimate of statistical uncertainty
- in the limit $N \rightarrow \infty$, MC estimate tends to normal distribution, uncertainty tends to standard deviation

Monte Carlo integration

- recap of Monte Carlo integration (non-uniform sampling):

$$\int_V p(\vec{x}) f(\vec{x}) d^D x \approx \langle f \rangle \pm \sqrt{\frac{\langle f^2 \rangle - \langle f \rangle^2}{N}}$$

$$\langle f \rangle \equiv \frac{1}{N} \sum_{i=1}^N f(\vec{x}_i) \qquad \langle f^2 \rangle \equiv \frac{1}{N} \sum_{i=1}^N f(\vec{x}_i)^2$$

- N points $\vec{x}_1, \dots, \vec{x}_N$ chosen **independently** and **randomly** with probability dist. $p(\vec{x})$ throughout D -dimensional volume V
- normalization condition $\int_V p(\vec{x}) d^D x = 1$
- law of large numbers justifies correctness of estimate
- central limit theorem gives estimate of statistical uncertainty

Pseudorandom number generators

- MC integration requires random numbers
- but computers are deterministic!!
- clever algorithms can produce sequences of numbers which **appear** to be random (pseudorandom)
 - uniform deviates between 0 and 1
- example: the **Mersenne twister**
 - <http://www.math.sci.hiroshima-u.ac.jp/~m-mat/MT/emt.html>
 - currently holds the record for longest period $2^{19937} - 1$
 - very fast, passes all standard tests (Diehard) for good RNG
- devising good RNGs is a science in itself
 - most utilize modulus function, bit shifting, shuffling

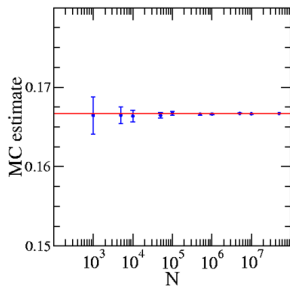
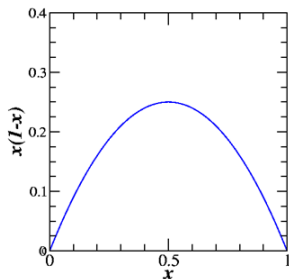


One-dimensional example

- simple example

$$\int_0^1 x(1-x) dx = \frac{1}{6} = 0.166666\dots$$

- plot of integrand and some Monte Carlo estimates



- not efficient for 1-dim integrals!

Importance sampling

- Monte Carlo method works best for flat functions, problems when integrand sharply peaked
- importance sampling can greatly improve efficiency of Monte Carlo integration → variance reduction
- recall simple integration

$$\int_a^b f(x) dx \approx \frac{(b-a)}{N} \sum_{j=1}^N f(x_j) \quad x_j \text{ chosen with uniform probability between } a \text{ and } b$$

- choose function $g(x) > 0$ with $\int_a^b g(x) dx = 1$ so $h(x) = \frac{f(x)}{g(x)}$ is as close as possible to a constant

$$\int_a^b f(x) dx = \int_a^b h(x)g(x) dx \approx \frac{(b-a)}{N} \sum_{j=1}^N h(x_j)$$

where x_j now chosen with probability density $g(x)$

- must be able to sample with probability density $g(x)$
- how to choose $g(\vec{x})$ for complicated multi-dimensional integral?

Sampling non-uniform distributions

- random number generators sample the uniform distribution
- to sample other densities, apply transformation method
- random variable U with uniform density $p_U(u) = 1$ for $0 \leq x \leq 1$
- transform to new random variable $Y = \phi(U)$ where ϕ is a strictly increasing function
 - strictly increasing function ensures inverse function is single-valued
 - also ensures that if $u + du > u$ then $y + dy > y$ for $y = \phi(u)$
- what is density p_Y ?
- from conservation of probability

$$p_Y(y)dy = p_U(u)du \quad p_Y(y) = p_U(u) \frac{du}{dy} = p_U(\phi^{-1}(y)) \frac{d\phi^{-1}(y)}{dy}$$

Sampling non-uniform distributions (continued)

- desired density p_Y is usually known, so must determine ϕ

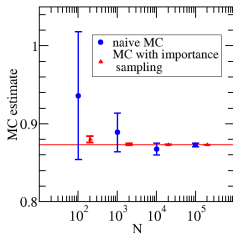
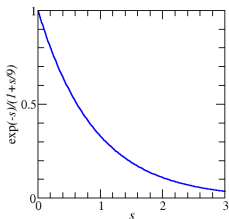
$$\int_0^u du' = \int_{\phi(0)}^{\phi(u)} p_Y(y) dy \Rightarrow u = F_Y(\phi(u)) \Rightarrow \phi(u) = F_Y^{-1}(u)$$

- F^{-1} unique since F is strictly increasing function
- summary: random variable Y with density $p_Y(y)$ and cumulative distribution $F_Y(y) = \int_{-\infty}^y p_Y(s) ds$ can be sampled by sampling with uniform deviate U then applying transformation

$$Y = F_Y^{-1}(U)$$

Exponential density

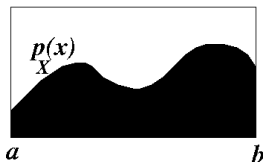
- transformation method requires density whose indefinite integral can be obtained and inverted
- method useful for only a handful of density functions
- one example: the exponential $p_Y(y) = e^{-y}/(1 - e^{-b})$ for $0 \leq y \leq b$
 - cumulative distribution $F_Y(y) = \int_0^y p_Y(s)ds = (1 - e^{-y})/(1 - e^{-b})$
 - inverse $F_Y^{-1}(u) = -\ln(1 - (1 - e^{-b})u)$
- example integral: $\int_0^3 \frac{e^{-s}}{1 + s/9} ds \approx 0.873109$



plot of integrand (left);
dramatic improvement
using importance
sampling (right)

Rejection method

- can sample from probability density whose cumulative distribution is not easily calculable and invertible using the **rejection method**
- sampling from density $p_X(x)$ for $a \leq x \leq b$
⇒ equivalent to choosing a random point in **two** dimensions with uniform probability in the **area** under curve $p_X(x)$
- simplest method: pick random point with uniform probability in box $a \leq x \leq b$ horizontally and $0 \leq y \leq \max(p_X(x))$ vertically
 - accept if below curve
 - reject if above curve, repeat until acceptance
- if $p_X(x)$ sharply peaked, use a comparison function $f(x)$ satisfying $f(x) \geq p_X(x)$ for all $a \leq x \leq b$ and which can be sampled by transformation method



Integrals of very high dimension

- sampling methods described so far work well in one-dimension
- for multi-dimensional integrals, transformation and rejection methods not feasible
- use of **stationary stochastic processes** to handle highly multi-dimensional integrals

Stochastic processes

- **stochastic process**: a sequence of events X_t , $t = 0, 1, 2, \dots$ governed by probabilistic laws (limit attention to discrete “time” t)
- consider a system which can be in one of R discrete states s_1, s_2, \dots, s_R (can generalize to continuum of states)
- system moves successively from one state to another
- each move is called a **step** (discrete “time”)
- given previous states of system X_0, X_1, \dots, X_{t-1} , conditional probability to find system in state X_t at time t is $P(X_0, \dots, X_{t-1} | X_t)$ which may depend on previous states of system and possibly t
- stochastic processes can be useful for Monte Carlo integration since sequence X_t samples conditional probability distribution $P(X_0, \dots, X_{t-1} | X_t)$

Stationary stochastic processes

- a stochastic process is **stationary** when the probabilistic laws remain unchanged through shifts in time
 - joint probability distribution of $(X_t, X_{t+j_1}, \dots, X_{t+j_n})$ is same as that of $(X_{t+h}, X_{t+h+j_1}, \dots, X_{t+h+j_n})$ for any h
- mean $E(X_t) = \mu$ is independent of t (if it exists)
- variance $E((X_t - \mu)^2) = \sigma^2$ independent of t if $E(X_t^2)$ finite
- now X_t are usually **not** independent random variables
- **autocovariance** $E((X_t - \mu)(X_s - \mu)) = R(|t - s|)$ depends only on time difference $|t - s|$
- define **autocorrelation function** $\rho(t) = R(t)/R(0)$ so that $\rho(0) = 1$ and $-1 \leq \rho(t) \leq 1$ for all t (from Schwartz's inequality)

Law of large numbers for stationary process

- consider a stationary process X_1, X_2, \dots with $E(X_k) = \mu$ and autocovariance $R(s) = E((X_k - \mu)(X_{k+s} - \mu))$ satisfying

$$\sum_{s=0}^{\infty} |R(s)| < \infty, \text{ and define } \bar{X}_N = \frac{1}{N}(X_1 + X_2 + \dots + X_N), \text{ then for}$$

$$\text{any } \varepsilon > 0, \quad \lim_{N \rightarrow \infty} P(|\bar{X}_N - \mu| \geq \varepsilon) = 0$$

- Proof:

- define $Y_n = X_n - \mu$ and $\bar{Y}_N = \frac{1}{N}(Y_1 + \dots + Y_N)$

$$\begin{aligned} E(\bar{Y}_N^2) &= \frac{1}{N^2} E\left(\sum_{k=1}^N Y_k^2 + 2 \sum_{k < l} Y_k Y_l\right) = \frac{1}{N^2} \left(NR(0) + 2 \sum_{k < l} R(l-k)\right) \\ &= \frac{1}{N^2} \left(2 \sum_{p=1}^N \sum_{h=0}^{p-1} R(h) - NR(0)\right) \\ &= \frac{1}{N} \left(R(0) + \sum_{k=1}^{N-1} 2R(k)(N-k)/N\right) \end{aligned}$$

Law of large numbers for stationary process (2)

- Proof (continued):

- continuing,

$$\begin{aligned} NE(\bar{Y}_N^2) &= \left| R(0) + \sum_{k=1}^{N-1} 2R(k)(N-k)/N \right| \\ &\leq |R(0)| + \sum_{k=1}^{N-1} 2|R(k)|(N-k)/N \\ &\leq |R(0)| + \sum_{k=1}^{N-1} 2|R(k)| \end{aligned}$$

- since $\sum_j |R(j)| < \infty$ then $NE(\bar{Y}_N^2) < \infty$ so $\lim_{N \rightarrow \infty} E(\bar{Y}_N^2) = 0$
- with Chebyshev inequality $P(|\bar{X}_N - \mu| \geq \varepsilon) \leq E((\bar{X}_N - \mu)^2)/\varepsilon^2$

$$\lim_{N \rightarrow \infty} E((\bar{X}_N - \mu)^2) = 0 \text{ implies } \lim_{N \rightarrow \infty} P(|\bar{X}_N - \mu| \geq \varepsilon) = 0$$

- limiting value $\lim_{N \rightarrow \infty} NE((\bar{X}_N - \mu)^2) = \sum_{k=-\infty}^{\infty} R(k)$

- Proof: given absolutely summable autocovariance $\sum_k |R(k)| < \infty$, for any $\varepsilon > 0$ there exists a q such that $\sum_{k=1}^{\infty} 2|R(q+k)| < \varepsilon/2$

- so $\left| \sum_{j=-(N-1)}^{N-1} R(j) - NE(\bar{Y}_N^2) \right| = \left| R(0) + 2 \sum_{j=1}^{\infty} R(j) - \left(R(0) + \sum_{k=1}^{N-1} 2R(k)(N-k)/N \right) \right|$

Law of large numbers for stationary process (3)

- limiting value $\lim_{N \rightarrow \infty} NE((\bar{X}_N - \mu)^2) = \sum_{k=-\infty}^{\infty} R(k)$

- Proof (continued):

$$\begin{aligned} \left| \sum_{j=-(N-1)}^{N-1} R(j) - NE(\bar{Y}_N^2) \right| &= \left| \sum_{k=1}^{N-1} 2kR(k)/N \right| \leq \sum_{k=1}^{N-1} 2k|R(k)| / N \\ &= \sum_{k=1}^q 2k|R(k)| / N + \sum_{k=q+1}^{N-1} 2k|R(k)| / N \\ &\leq \sum_{k=1}^q 2k|R(k)| / N + \sum_{k=q+1}^{N-1} 2|R(k)| \\ &\leq \sum_{k=1}^q 2k|R(k)| / N + \varepsilon/2 \end{aligned}$$

- since q fixed and finite, can always increase N so that $\sum_{k=1}^q 2k|R(k)| / N < \varepsilon/2$ which holds as $N \rightarrow \infty$
- so $\left| \sum_{j=-(N-1)}^{N-1} R(j) - NE(\bar{Y}_N^2) \right| < \varepsilon$ which proves the limit

M -dependent central limit theorem

- M -dependent central limit theorem: Let X_1, X_2, \dots, X_N be a stationary M -dependent sequence of random variables (X_t and X_{t+s} are independent for $s > M$) such that $E(X_t) = E(X_1) = \mu$ and $E((X_1 - \mu)^2) < \infty$, and define $\bar{X}_N = (X_1 + X_2 + \dots + X_N)/N$ and $\sigma^2 = E((X_1 - \mu)^2) + 2 \sum_{h=1}^M E((X_1 - \mu)(X_{h+1} - \mu))$. Then for $a < b$,

$$\lim_{N \rightarrow \infty} P \left(\frac{a\sigma}{\sqrt{N}} < (\bar{X}_N - \mu) < \frac{b\sigma}{\sqrt{N}} \right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx$$

- distribution of $(X_1 + \dots + X_N - N\mu)/(\sigma\sqrt{N})$ tends to standard normal (zero mean, unit variance)
- for proof, see W. Hoeffding and H. Robbins, Duke Math. Journal **15**, 773 (1948) or T. Anderson, The Statistical Analysis of Time Series, Wiley (1971).
- autocovariance $R(h) = R(-h) = E((X_t - \mu)(X_{t+|h|} - \mu))$
- note that $\sigma^2 = \sum_{h=-M}^M R(h) = NE((\bar{X}_N - \mu)^2)$ for $N \gg M$

Monte Carlo integration

- Monte Carlo integration using stationary stochastic process:

$$\int_V p(\vec{x}) f(\vec{x}) d^D x \approx \langle f \rangle \pm \sqrt{\frac{R_0(f) + 2 \sum_{h \geq 1} R_h(f)}{N}}$$
$$\langle f \rangle \equiv \frac{1}{N} \sum_{i=1}^N f(\vec{x}_i), \quad R_h(f) \equiv \frac{1}{N-h} \sum_{i=1}^{N-h} (f(\vec{x}_i) - \langle f \rangle) (f(\vec{x}_{i+h}) - \langle f \rangle)$$

- N points $\vec{x}_1, \dots, \vec{x}_N$ are **stationary** sequence of **random** variables with probability dist. $p(\vec{x})$ throughout D -dimensional volume V
- normalization condition $\int_V p(\vec{x}) d^D x = 1$
- absolutely summable autocovariance $\sum_{h=0}^{\infty} |R_h(f)| < \infty$
- law of large numbers justifies correctness of estimate
- m -dependent central limit theorem gives estimate of statistical uncertainty

Simple stationary stochastic processes

- how can we find a stationary stochastic process that yields the desired probability distribution $p(\vec{x})$?
- use of **Markov chains** to solve this problem



A. A. Markov (1886).

- Markov chains were introduced by the Russian mathematician Andrei Markov (1856-1922) in 1906

Markov chains

- discrete Markov chain: stochastic process which generates a sequence of states with probabilities depending only on current state
 - consider a system which can be in one of R states s_1, s_2, \dots, s_R
 - system moves successively from one state to another
 - each move is called a **step** (discrete “time”)
 - if current state is s_i , then chain moves to state s_j at next step with probability p_{ij} which does **not** depend on previous states of chain
 - probabilities p_{ij} are called **transition probabilities**
 - the square $R \times R$ real-valued matrix \mathbf{P} whose elements are p_{ij} is called the **transition matrix** or the **Markov matrix**
- **time homogeneous** if transition probabilities p_{ij} independent of “time” or position in chain
- definition generalizes to continuous set of states
 - leads to matrix of transition densities
- will not deal with continuous-time chains here

Some basic properties of Markov chains

- transition matrix \mathbf{P} has non-negative entries $p_{ij} \geq 0$
- since probability of going from s_i to **any** state must be unity, then matrix elements must satisfy $\sum_{j=1}^R p_{ij} = 1$ (rows sum to unity)
- if columns also sum to unity, \mathbf{P} is called **doubly stochastic** matrix
- if \mathbf{P}_1 and \mathbf{P}_2 are Markov matrices, then the matrix product $\mathbf{P}_1\mathbf{P}_2$ is also a Markov matrix
- every eigenvalue λ of a Markov matrix satisfies $|\lambda| \leq 1$
- every Markov matrix has at least one eigenvalue equal to unity

Eigenvalues/eigenvectors of real square matrices

- for a square matrix \mathbf{P} , a nonzero column vector \mathbf{v} which satisfies $\mathbf{P}\mathbf{v} = \lambda\mathbf{v}$ for complex scalar λ is known as a **right eigenvector** corresponding to **eigenvalue** λ
 - often, “right eigenvectors” are simply called “eigenvectors”
- a nonzero vector \mathbf{v} satisfying $\mathbf{v}^T\mathbf{P} = \lambda\mathbf{v}^T$, where T indicates **transpose**, is known as a **left eigenvector**
- every square $R \times R$ matrix has R complex eigenvalues, counting multiple roots according to their multiplicity
- for a real square matrix, the eigenvalues are either real or come in complex conjugate pairs
- eigenvectors for distinct eigenvalues are linearly independent
- a degenerate eigenvalue may not have distinct eigenvectors
- R linearly independent eigenvectors guaranteed only if all R eigenvalues distinct
- a matrix \mathbf{P} and its transpose \mathbf{P}^T have the same eigenvalues

Properties of Markov matrices (continued)

- every eigenvalue λ of Markov matrix \mathbf{P} satisfies $|\lambda| \leq 1$
 - Proof: suppose complex number λ is an eigenvalue of \mathbf{P} with corresponding eigenvector \mathbf{v} so that $\mathbf{P}\mathbf{v} = \lambda\mathbf{v}$
 - let k be such that $|v_k| \geq |v_j|$ for all j
 - k -th component of eigenvalue equation gives us $\sum_j p_{kj}v_j = \lambda v_k$
 - use generalized triangle inequality for complex numbers $|\sum_k z_k| \leq \sum_k |z_k|$ to show $|\lambda v_k| = |\sum_j p_{kj}v_j| \leq \sum_j p_{kj}|v_j| \leq \sum_j p_{kj}|v_k| = |v_k|$
 - thus, $|\lambda v_k| = |\lambda||v_k| \leq |v_k| \rightarrow |\lambda| \leq 1$
- every Markov matrix \mathbf{P} has a least one eigenvalue equal to unity
 - Proof: let \mathbf{v} be a vector satisfying $v_j = 1$ for all j
 - then $\sum_j p_{ij}v_j = \sum_j p_{ij} = 1 = v_i$
 - hence, \mathbf{v} is an eigenvector corresponding to eigenvalue 1

Multi-step probabilities

- **n -step transition probability**: ij -th element $p_{ij}^{(n)}$ of matrix \mathbf{P}^n is probability that Markov chain, starting in state s_i , will be in state s_j after n steps
 - probability to go from s_i to s_j in 2 steps is $\sum_{k=1}^R p_{ik}p_{kj}$
 - generalizes to n -steps
- for starting probability vector \mathbf{u} , probability that chain in state s_j after n steps is $u_j^{(n)} = \sum_{i=1}^R u_i p_{ij}^{(n)}$
 - u_i is probability starting state is s_i , matrix form $\mathbf{u}^{(n)T} = \mathbf{u}^T \mathbf{P}^n$
- **first visit probability**: the probability that a Markov chain, starting in state s_i , is found for the **first** time in state s_j after n steps \rightarrow denoted by $f_{ij}^{(n)}$
 - define $f_{ij}^{(0)} = 0$ one step, $f_{ij}^{(1)} = p_{ij}$, two steps, $f_{ij}^{(2)} = \sum_{k \neq j} p_{ik}p_{kj}$
 - generalize $f_{ij}^{(n)} = \sum_{k \neq j} p_{ik} f_{kj}^{(n-1)}$
- important relation for later user: $p_{ij}^{(n)} = \sum_{m=1}^n f_{ij}^{(m)} p_{jj}^{(n-m)}$

Mean first passage and mean recurrence times

- **total visit probability**: probability that, starting from state s_i , chain will **ever** visit state s_j :

$$f_{ij} = \sum_{n=1}^{\infty} f_{ij}^{(n)}$$

- **mean first passage time** from s_i to s_j is expected number of steps to reach state s_j in an ergodic Markov chain for the first time, starting from state $s_i \rightarrow$ denoted by m_{ij} (by convention, $m_{ii} = 0$)

$$m_{ij} = \sum_{n=1}^{\infty} n f_{ij}^{(n)}$$

- **mean recurrence time** μ_i of state s_i is expected number of steps to return to state s_i for the first time in an ergodic Markov chain starting from s_i

$$\mu_i = \sum_{n=1}^{\infty} n f_{ii}^{(n)}$$

Classes

- state s_j is **accessible** from state s_i if $p_{ij}^{(n)} > 0$ for some finite n
 - often denoted by $s_i \rightarrow s_j$
 - if $s_i \rightarrow s_j$ and $s_j \rightarrow s_k$, then $s_i \rightarrow s_k$
- states s_i and s_j **communicate** if $s_i \rightarrow s_j$ and $s_j \rightarrow s_i$
 - denoted by $s_i \leftrightarrow s_j$
 - $s_i \leftrightarrow s_j$ and $s_j \leftrightarrow s_k$ implies $s_i \leftrightarrow s_k$
- **class** = a set of states that all communicate with one another
- if C_1 and C_2 are communicating classes, then either $C_1 = C_2$ or C_1, C_2 are disjoint
 - if C_1 and C_2 have a common state s_i , then $s_i \leftrightarrow s_{j1}$ for all $s_{j1} \in C_1$ and $s_i \leftrightarrow s_{j2}$ for all $s_{j2} \in C_2$, so $s_{j1} \leftrightarrow s_{j2}$ implying $C_1 = C_2$
- set of all states can be partitioned into separate classes
 - if transition from class C_1 to different class C_2 is possible, then transition from C_2 to C_1 not possible, otherwise $C_1 = C_2$

Irreducible Markov chains

- a Markov chain is called **irreducible** if the probability to go from every state to every state (not necessarily in one step) is greater than zero
- all states in irreducible chain are in one single communicating class

Classification of states in Markov chains

- states in a Markov chain are
 - (a) **recurrent** (persistent) or **transient**
 - recurrent states are either **positive** or **null**
 - (b) **periodic** (cyclic) or **aperiodic**
- **recurrent** or **persistent** state has $f_{ii} = \sum_{n=1}^{\infty} f_{ii}^{(n)} = 1$
 - unit probability of returning to state after a finite length
- **transient** state has $f_{ii} = \sum_{n=1}^{\infty} f_{ii}^{(n)} < 1$
- recurrent state is **positive** if mean recurrence time finite $\mu_i < \infty$
otherwise, called **null**
- the **period** of a state in a Markov chain is the greatest common divisor of all $n \geq 0$ for which $p_{ii}^{(n)} > 0$
 - transition s_i to s_i not possible except for multiples of period $d(i)$
- **periodic** state s_i has period $d(i) > 1$
aperiodic state s_i has period $d(i) = 1$

Recurrent and transient states

- for a recurrent state, $\sum_{n=1}^{\infty} p_{ii}^{(n)} = \infty$, whereas
for a transient state, $\sum_{n=1}^{\infty} p_{ii}^{(n)} < \infty$
- proof:
 - we start with the following:

$$\sum_{n=1}^N p_{ij}^{(n)} = \sum_{n=1}^N \sum_{m=1}^n f_{ij}^{(m)} p_{jj}^{(n-m)} = \sum_{m=1}^N f_{ij}^{(m)} \sum_{n=0}^{N-m} p_{jj}^{(n)} \leq \sum_{m=1}^N f_{ij}^{(m)} \sum_{n=0}^N p_{jj}^{(n)}$$

- but for $N > N'$ we also have

$$\sum_{n=1}^N p_{ij}^{(n)} = \sum_{m=1}^N f_{ij}^{(m)} \sum_{n=0}^{N-m} p_{jj}^{(n)} \geq \sum_{m=1}^{N'} f_{ij}^{(m)} \sum_{n=0}^{N-m} p_{jj}^{(n)} \geq \sum_{m=1}^{N'} f_{ij}^{(m)} \sum_{n=0}^{N-N'} p_{jj}^{(n)}$$

- putting together above results:

$$\sum_{m=1}^{N'} f_{ij}^{(m)} \sum_{n=0}^{N-N'} p_{jj}^{(n)} \leq \sum_{n=1}^N p_{ij}^{(n)} \leq \sum_{m=1}^N f_{ij}^{(m)} \sum_{n=0}^N p_{jj}^{(n)}$$

- take $N \rightarrow \infty$ first, then $N' \rightarrow \infty$ to get

$$f_{ij} \sum_{n=0}^{\infty} p_{jj}^{(n)} \leq \sum_{n=1}^{\infty} p_{ij}^{(n)} \leq f_{ij} \sum_{n=0}^{\infty} p_{jj}^{(n)} \Rightarrow f_{ij} \sum_{n=0}^{\infty} p_{jj}^{(n)} = \sum_{n=1}^{\infty} p_{ij}^{(n)}$$

Recurrent and transient states (2)

- for a recurrent state, $\sum_{n=1}^{\infty} p_{ii}^{(n)} = \infty$, whereas
for a transient state, $\sum_{n=1}^{\infty} p_{ii}^{(n)} < \infty$

- proof (continued):

- so far have shown $f_{ij} \sum_{n=0}^{\infty} p_{jj}^{(n)} = \sum_{n=1}^{\infty} p_{ij}^{(n)}$

- set $i = j$ then $f_{ii}(1 + \sum_{n=1}^{\infty} p_{ii}^{(n)}) = \sum_{n=1}^{\infty} p_{ii}^{(n)}$

- so finally

$$\sum_{n=1}^{\infty} p_{ii}^{(n)} = \frac{f_{ii}}{1 - f_{ii}}$$

- $f_{ii} = 1$ for a recurrent state and $f_{ii} < 1$ for a transient state, which proves the above statements

- note that the above results also imply

$$\sum_{n=1}^{\infty} p_{ij}^{(n)} = \frac{f_{ij}}{1 - f_{ii}}$$

Recurrent and transient states (3)

- a Markov chain returns to a **recurrent** state infinitely often and returns to a **transient** state only a finite number of times
- proof:
 - let $g_{ij}(m)$ denote probability that chain enters state s_j at least m times, starting from s_i
 - clearly $g_{ij}(1) = f_{ij}$
 - one also sees $g_{ij}(m+1) = f_{ij}g_{jj}(m)$ so $g_{ij}(m) = (f_{ij})^m$
 - probability of entering s_j infinitely many times is $g_{ij} = \lim_{m \rightarrow \infty} g_{ij}(m) = \lim_{m \rightarrow \infty} (f_{ij})^m$
 - so starting in s_j then

$$g_{jj} = \lim_{m \rightarrow \infty} (f_{jj})^m = \begin{cases} 1 & \text{for recurrent state } f_{jj} = 1 \\ 0 & \text{for transient state } f_{jj} < 1 \end{cases}$$

Important result for recurrent states

- if s_i is recurrent and $s_i \rightarrow s_j$, then $f_{ji} = 1$
- proof:
 - let $\alpha > 0$ denote probability to reach s_j from s_i without previously returning to s_i
 - probability of **never** returning to s_i from s_j is $1 - f_{ji}$
 - probability of never returning to s_i from s_i is at least $\alpha(1 - f_{ji})$
 - but s_i is recurrent so probability of no return is zero
 - thus, $f_{ji} = 1$
- for two communicating states $s_i \leftrightarrow s_j$ that are each recurrent, it follows that $f_{ij} = f_{ji} = 1$

Similarity of states in a class

- all states in a class of a Markov chain are of the same type, and if periodic, all have the same period

- proof:

- for any two states s_i and s_j in a class, there exists integers r and s such that $p_{ij}^{(r)} = \alpha > 0$ and $p_{ji}^{(s)} = \beta > 0$ so

$$p_{ii}^{(n+r+s)} = \sum_{kl} p_{ik}^{(r)} p_{kl}^{(n)} p_{li}^{(s)} \geq \sum_k p_{ik}^{(r)} p_{kk}^{(n)} p_{ki}^{(s)} \geq p_{ij}^{(r)} p_{jj}^{(n)} p_{ji}^{(s)} = \alpha \beta p_{jj}^{(n)}$$

- if s_i is transient, then left-hand side is a term of a convergent series $\sum_k p_{ii}^{(k)} < \infty$, so the same must be true for $p_{jj}^{(k)}$, and if $p_{ii}^{(k)} \rightarrow 0$, then $p_{jj}^{(k)} \rightarrow 0$
- the same statements remain true if the roles of i and j are reversed, so either both s_i and s_j are transient, or neither is
- if s_j is null (infinite mean recurrence time $\mu_j = \sum_{n=1}^{\infty} n f_{jj}^{(n)} = \infty$), then s_i must be null as well
- same statements true if i, j reversed, so if one is a null state, then so is the other

Similarity of states in a class (2)

- again, we have

- for any two states s_i and s_j in a class, there exists integers r and s such that $p_{ij}^{(r)} = \alpha > 0$ and $p_{ji}^{(s)} = \beta > 0$ so

$$p_{ii}^{(n+r+s)} = \sum_{kl} p_{ik}^{(r)} p_{kl}^{(n)} p_{li}^{(s)} \geq \sum_k p_{ik}^{(r)} p_{kk}^{(n)} p_{ki}^{(s)} \geq p_{ij}^{(r)} p_{jj}^{(n)} p_{ji}^{(s)} = \alpha\beta p_{jj}^{(n)}$$

- suppose s_i has period t , then for $n = 0$, the right-hand side is positive, so $p_{ii}^{(r+s)} > 0$ which means that $r + s$ must be a multiple of t
- hence, left-hand side vanishes unless n is multiple of t , so $p_{jj}^{(n)}$ can be nonzero only if n is multiple of t , so s_i and s_j have same period
- chain aperiodic if $p_{ii} > 0$ for at least one s_i

Periodic irreducible chains

- states in an irreducible chain with period d can be partitioned into d mutually exclusive subsets G_0, \dots, G_{d-1} such that if state $s_k \in G_\alpha$, then $p_{1k}^{(n)} = 0$ unless $n = \alpha + \nu d$
- proof:
 - since irreducible, all states have same period d and every state can be reached from every other state
 - there exist for every state s_k two integers a and b such that $p_{1k}^{(a)} > 0$ and $p_{1k}^{(b)} > 0$
 - but $p_{11}^{(a+b)} = \sum_j p_{1j}^{(a)} p_{j1}^{(b)} \geq p_{1k}^{(a)} p_{k1}^{(b)} > 0$ so $a + b$ divisible by d
 - thus, $a + b = md$ for integer m , or $a = -b + md$
 - rewrite as $a = \alpha + \nu d$ for integer ν and $0 \leq \alpha < d$
 - α is characteristic of state s_k so all states partitioned into d mutually exclusive subsets G_0, G_1, \dots, G_{d-1}
- with proper ordering of G_α subsets, **one-step** transition from state in G_α always leads to state in $G_{\alpha+1}$, or from G_{d-1} to G_0
- each subset G_α closed in aperiodic Markov chain with transition matrix \mathbf{P}^d

Fact concerning finite Markov chains

- in an irreducible chain having finite number R of states, there are no null states and it is impossible that all states are transient
- proof:
 - all rows of the matrix \mathbf{P}^n must add to unity
 - since each row contains finite number of non-negative elements, it is impossible that $p_{ij}^{(n)} \rightarrow 0$ for all i, j pairs
 - thus, impossible that all states are transient
 - so at least one state must be non-null
 - but since irreducible (one class), all states must be non-null
- in an R -state irreducible Markov chain, it is possible to go from any state to any other state in at most $R - 1$ steps

A crucial theorem about two sequences

- important theorem: (**basic limit theorem** of the **renewal equation**)
given a sequence f_0, f_1, f_2, \dots such that

$$f_0 = 0, \quad f_n \geq 0, \quad \sum_{n=0}^{\infty} f_n = 1$$

and greatest common divisor of those n for which $f_n > 0$ is $d \geq 1$
and another sequence u_0, u_1, u_2, \dots defined by

$$u_0 = 1, \quad u_n = \sum_{m=1}^n f_m u_{n-m} \quad (n \geq 1)$$

then

$$\lim_{n \rightarrow \infty} u_{nd} = \begin{cases} d\mu^{-1} & \text{if } \mu = \sum_{n=1}^{\infty} n f_n < \infty \\ 0 & \text{if } \mu = \infty \end{cases}$$

- proof:
 - see W. Feller, *An Introduction to Probability Theory and Its Applications*, Vol. I.
 - or S. Karlin and H. Taylor, *A First Course in Stochastic Processes*.

Basic limit theorem

- we shall only sketch the proof of this theorem
- first, some key properties of these sequences
 - $0 \leq f_n \leq 1$ for all n since $f_n \geq 0$ and $\sum_{n=0}^{\infty} f_n = 1$
 - $0 \leq u_n \leq 1$ for all n can be established inductively
 - $u_0 = 1, u_1 = f_1, u_2 = f_2 + f_1^2$ satisfy above bounds
 - assume $0 \leq u_k \leq 1$ for all $0 \leq k \leq n$
 - since $f_m \geq 0$ and $\sum_{m=1}^{\infty} f_m = 1$ then
$$u_{n+1} = \sum_{m=1}^{n+1} f_m u_{n+1-m} \geq 0$$
 since sum of nonnegative terms, and
$$u_{n+1} = \sum_{m=1}^{n+1} f_m u_{n+1-m} \leq \sum_{m=1}^{n+1} f_m \leq 1$$
- next, limit our attention to $d = 1$ (nonperiodic)
- since u_n is a bounded sequence, $\lambda \equiv \limsup_{n \rightarrow \infty} u_n$ is finite and there exists a subsequence $n_1 < n_2 < \dots$ tending to infinity such that $\lim_{j \rightarrow \infty} u_{n_j} = \lambda$
- next step in proof is to show $\lim_{j \rightarrow \infty} u_{n_j - q} = \lambda$ for any integer $q \geq 0$ when $f_1 > 0$ (we'll skip this)

Basic limit theorem (2)

- define a new sequence $r_n = \sum_{k>n} f_k$
- some important properties of this sequence
 - $r_n \geq 0$ for all n , and $r_0 = 1$
 - $r_{n-1} - r_n = f_n$ for $n \geq 1$
 - $\sum_{n=0}^{\infty} r_n = \sum_{n=1}^{\infty} n f_n \equiv \mu$
- one very crucial identity: $\sum_{k=0}^N r_k u_{N-k} = 1$ for all $N \geq 0$
 - define $A_N = \sum_{k=0}^N r_k u_{N-k}$
 - start with $u_N = \sum_{m=1}^N f_m u_{N-m} = \sum_{m=1}^N (r_{m-1} - r_m) u_{N-m}$
 - use $r_0 = 1$ and rearrange $r_0 u_N + \sum_{m=1}^N r_m u_{N-m} = \sum_{m=1}^N r_{m-1} u_{N-m}$
 - take $m \rightarrow k+1$ on right: $\sum_{m=0}^N r_m u_{N-m} = \sum_{k=0}^{N-1} r_k u_{N-1-k}$
 - have shown $A_N = A_{N-1}$ for all N
 - $A_N = A_{N-1} = A_{N-2} = \dots = A_0 = r_0 u_0 = 1$

Basic limit theorem (3)

- recall that $n_1 < n_2 < \dots$ is subsequence such that $\lim_{j \rightarrow \infty} u_{n_j - q} = \lambda$ for any integer $q \geq 0$
- since $\sum_{k=0}^{n_j} r_k u_{n_j - k} = 1$ for all n_j and $r_k \geq 0$, $u_k \geq 0$ for all k , then $\sum_{k=0}^N r_k u_{n_j - k} \leq 1$ for fixed $N < n_j$
- take limit $j \rightarrow \infty$ so $\lim_{j \rightarrow \infty} \sum_{k=0}^N r_k u_{n_j - k} = \lambda \sum_{k=0}^N r_k \leq 1$
- already know $\lambda \geq 0$, take $N \rightarrow \infty$ to have $0 \leq \lambda \leq 1 / (\sum_{k=0}^{\infty} r_k)$
- if $\sum_{k=0}^{\infty} r_k = \infty$ then $\lim_{n \rightarrow \infty} u_n = \lambda = 0$
- if $\mu = \sum_{k=0}^{\infty} r_k$ is finite, $N \rightarrow \infty$ gives $\mu \lambda \leq 1$
- define $M = \sup_{n \geq 0} u_n$ so $0 \leq u_k \leq M \leq 1$ for all k
- define $g(D) = \sum_{k=D+1}^{\infty} r_k$, note $g(D) \geq 0$ for all D and $\lim_{D \rightarrow \infty} g(D) = 0$
- consider $\sum_{k=0}^D r_k u_{n_j - k} + \sum_{k=D+1}^{n_j} r_k u_{n_j - k} = 1$ for $D < n_j$
- thus $\sum_{k=0}^D r_k u_{n_j - k} + M g(D) \geq 1$ for $D < n_j$

Basic limit theorem (4)

- again, $\sum_{k=0}^D r_k u_{n_j-k} + Mg(D) \geq 1$ for $D < n_j$
- take $j \rightarrow \infty$ to conclude $\lambda \left(\sum_{k=0}^D r_k \right) + Mg(D) \geq 1$
- take limit $D \rightarrow \infty$ to obtain $\lambda\mu \geq 1$
- have now shown $1 \leq \mu\lambda \leq 1$ so $\mu\lambda = 1$
- proof for nonperiodic ($d = 1$) case now complete
- when $d > 1$ then $f_m = 0$ unless $m = nd$
- can then show $u_m = 0$ unless $m = nd$
- define new sequences $f'_n = f_{nd}$ and $u'_n = u_{nd}$ for $n = 0, 1, 2, \dots$
- since new sequence aperiodic, know $\lim_{n \rightarrow \infty} u'_n = 1/\mu'$ where $\mu' = \sum_{n=0}^{\infty} n f'_n$
- since $f_m = 0$ when $m \neq nd$ then $\mu' = \sum_{n=0}^{\infty} n f_{nd} = d^{-1} \sum_{m=0}^{\infty} m f_m = \mu/d$
- thus, $\lim_{n \rightarrow \infty} u_{nd} = d\mu^{-1}$ as required

Asymptotic behavior of $p_{jj}^{(n)}$

- asymptotic behavior of $p_{jj}^{(n)}$ can be summarized as

$$\lim_{n \rightarrow \infty} p_{jj}^{(dn)} = \begin{cases} 0 & s_j \text{ transient or null recurrent} \\ \mu_j^{-1} & s_j \text{ aperiodic positive recurrent} \\ d\mu_j^{-1} & s_j \text{ positive recurrent with period } d \end{cases}$$

- proof:

- if s_j transient, $\sum_n p_{jj}^{(n)}$ finite (converges) requiring $p_{jj}^{(n)} \rightarrow 0$
- for recurrent s_j , let $f_n = f_{jj}^{(n)}$ and $u_n = p_{jj}^{(n)}$
- sequences f_n, u_n so defined satisfy conditions of basic limit theorem
- basic limit theorem gives $p_{jj}^{(dn)} \rightarrow d\mu_j^{-1}$ where $\mu_j = \sum_n n f_{jj}^{(n)}$ is mean recurrence time
- aperiodic case when $d = 1$
- null recurrent s_j has $\mu_j = \infty$ so $p_{jj}^{(n)} \rightarrow \mu_j^{-1} = 0$

Asymptotic behavior of $p_{ij}^{(n)}$

- asymptotic behavior of $p_{ij}^{(n)}$ can be summarized as

$$\lim_{n \rightarrow \infty} p_{ij}^{(n)} = \begin{cases} 0 & s_j \text{ transient or null recurrent} \\ f_{ij} \mu_j^{-1} & s_j \text{ aperiodic positive recurrent} \end{cases}$$

- ignore periodic case here
- proof:

- $$p_{ij}^{(n)} = \sum_{m=1}^n f_{ij}^{(m)} p_{jj}^{(n-m)} = \sum_{m=1}^{n'} f_{ij}^{(m)} p_{jj}^{(n-m)} + \sum_{m=n'+1}^n f_{ij}^{(m)} p_{jj}^{(n-m)} \quad (n' < n)$$

- since $0 \leq \sum_{m=n'+1}^n f_{ij}^{(m)} p_{jj}^{(n-m)} \leq \sum_{m=n'+1}^n f_{ij}^{(m)}$ then

$$0 \leq \left(p_{ij}^{(n)} - \sum_{m=1}^{n'} f_{ij}^{(m)} p_{jj}^{(n-m)} \right) \leq \sum_{m=n'+1}^n f_{ij}^{(m)} \quad (n' < n)$$

- take $n \rightarrow \infty$, then $n' \rightarrow \infty$ above, denote $p_{jj} = \lim_{n \rightarrow \infty} p_{jj}^{(n)}$

$$0 \leq \left(\lim_{n \rightarrow \infty} p_{ij}^{(n)} - p_{ij} f_{ij} \right) \leq 0 \quad \Rightarrow \quad \lim_{n \rightarrow \infty} p_{ij}^{(n)} = p_{ij} f_{ij}$$

- for s_j transient or null recurrent, $p_{jj} = 0$ and f_{ij} finite, so $\lim_{n \rightarrow \infty} p_{ij}^{(n)} = 0$

- for s_j aperiod positive recurrent, $p_{jj} = \mu_j^{-1}$ so $p_{ij}^{(n)} \rightarrow f_{ij} \mu_j^{-1}$

Fixed-point or stationary distributions

- a probability vector \mathbf{w} is called **stationary** or **invariant** or a **fixed-point** if $\mathbf{w}^T = \mathbf{w}^T \mathbf{P}$
- clearly, one also has $\mathbf{w}^T = \mathbf{w}^T \mathbf{P}^n$
- the probability vector is always the same (stationary) for the chain
- when this occurs, the Markov chain is said to be in **equilibrium**

Fatou's lemma

- lemma: let $a_n(t)$ for $n = 1, 2, \dots$ be a function on a discrete set $T = \{1, 2, \dots\}$, assume $\lim_{n \rightarrow \infty} a_n(t)$ exists for each t in T , and suppose $a_n(t) \geq 0$ for all t, n , then

$$\sum_{t \in T} \left(\lim_{n \rightarrow \infty} a_n(t) \right) \leq \lim_{n \rightarrow \infty} \sum_{t \in T} a_n(t)$$

- proof:

- for any integer M

$$\sum_{t=1}^M \left(\lim_{n \rightarrow \infty} a_n(t) \right) = \lim_{n \rightarrow \infty} \sum_{t=1}^M a_n(t) \leq \lim_{n \rightarrow \infty} \sum_{t=1}^{\infty} a_n(t)$$

since all $a_n(t) \geq 0$

- take limit $M \rightarrow \infty$ to obtain required result

- example: $a_n(t) = \frac{n}{n^2 + t^2}$

- for $n > t$ then $\lim_{n \rightarrow \infty} a_n(t) = 0$ so $\sum_{t=1}^{\infty} \left(\lim_{n \rightarrow \infty} a_n(t) \right) = 0$

- $\sum_{t=1}^{\infty} a_n(t) = \frac{\pi}{2} \coth(n\pi) - \frac{1}{2n}$ so $\lim_{n \rightarrow \infty} \sum_{t=1}^{\infty} a_n(t) = \frac{\pi}{2}$

Dominated convergence theorem

- theorem: let $a_n(t)$ for $n = 1, 2, \dots$ be a function on a discrete set $T = \{1, 2, \dots\}$, assume $\lim_{n \rightarrow \infty} a_n(t)$ exists for each t in T , and suppose a function $B(t)$ exists such that $|a_n(t)| \leq B(t)$ for all t, n and $\sum_{t \in T} B(t) < \infty$, then

$$\sum_{t \in T} \left(\lim_{n \rightarrow \infty} a_n(t) \right) = \lim_{n \rightarrow \infty} \sum_{t \in T} a_n(t)$$

- proof:

- let $a(t) = \lim_{n \rightarrow \infty} a_n(t)$ and since $|a(t)| \leq B(t)$ then $\sum_{t=1}^{\infty} a(t)$ converges

- for any integer M

$$\left| \sum_{t=1}^{\infty} a_n(t) - \sum_{t=1}^{\infty} a(t) \right| \leq \sum_{t=1}^M |a_n(t) - a(t)| + \sum_{t=M+1}^{\infty} (|a_n(t)| + |a(t)|)$$

- now $\lim_{n \rightarrow \infty} \sum_{t=1}^M |a_n(t) - a(t)| = \sum_{t=1}^M \left(\lim_{n \rightarrow \infty} |a_n(t) - a(t)| \right) = 0$
 $\sum_{t=M+1}^{\infty} (|a_n(t)| + |a(t)|) \leq 2 \sum_{t=M+1}^{\infty} B(t)$

- so for any integer M

$$\left| \lim_{n \rightarrow \infty} \sum_{t=1}^{\infty} a_n(t) - \sum_{t=1}^{\infty} \lim_{n \rightarrow \infty} a_n(t) \right| \leq 2 \sum_{t=M+1}^{\infty} B(t)$$

- right-hand side is remainder of convergent series so equals zero in $M \rightarrow \infty$ limit

Fundamental limit theorem for ergodic Markov chains

- Theorem: an **irreducible aperiodic** Markov chain with transition matrix \mathbf{P} has a **stationary** distribution \mathbf{w} satisfying $w_j > 0$, $\sum_j w_j = 1$, and $\mathbf{w}^T = \mathbf{w}^T \mathbf{P}$ if, and only if, all its states are **positive recurrent**, and this stationary distribution is **unique** and identical to the limiting distribution $w_j = \lim_{n \rightarrow \infty} p_{ij}^{(n)}$ independent of initial state s_i
- Proof:
 - for irreducible aperiodic chain, the following possibilities exist:
 - (a) **all** states are positive recurrent
 - (b) **all** states are null recurrent
 - (c) **all** states are transient
 - if all states transient or null recurrent, $\lim_{n \rightarrow \infty} p_{ij}^{(n)} = 0$
 - if all states positive recurrent, then since **all** states communicate, $f_{ij} = 1$ for all i, j and previous result becomes $\lim_{n \rightarrow \infty} p_{ij}^{(n)} = \mu_j^{-1}$
 - can define $w_j = \lim_{n \rightarrow \infty} p_{ij}^{(n)}$ which is independent of initial state s_i
 - for all states positive recurrent, then $0 < \mu_j < \infty$ so $w_j > 0$ for all j

Fundamental limit theorem (2)

- Theorem: an **irreducible aperiodic** Markov chain with transition matrix \mathbf{P} has a **stationary** distribution \mathbf{w} satisfying $w_j > 0$, $\sum_j w_j = 1$, and $\mathbf{w}^T = \mathbf{w}^T \mathbf{P}$ if, and only if, all its states are **positive recurrent**, and this stationary distribution is **unique** and identical to the limiting distribution $w_j = \lim_{n \rightarrow \infty} p_{ij}^{(n)}$ independent of initial state s_i
- Proof (continued):

- we have $p_{ij}^{(m+n)} = \sum_{k=1}^{\infty} p_{ik}^{(n)} p_{kj}^{(m)}$ so using Fatou's lemma:

$$\lim_{n \rightarrow \infty} p_{ij}^{(m+n)} = \lim_{n \rightarrow \infty} \sum_{k=1}^{\infty} p_{ik}^{(n)} p_{kj}^{(m)} \geq \sum_{k=1}^{\infty} \lim_{n \rightarrow \infty} p_{ik}^{(n)} p_{kj}^{(m)}$$

- taking the limit $n \rightarrow \infty$ yields $w_j \geq \sum_{k=1}^{\infty} w_k p_{kj}^{(m)}$
- define $s \equiv \sum_{k=1}^{\infty} w_k$ then sum above equation over j :

$$s = \sum_{j=1}^{\infty} w_j \geq \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} w_k p_{kj}^{(m)} = \sum_{k=1}^{\infty} w_k \sum_{j=1}^{\infty} p_{kj}^{(m)} = \sum_{k=1}^{\infty} w_k = s$$

interchanging order of the two infinite summations is possible since all summands non-negative (Fubini's theorem)

- since $s \geq s$, equality must hold for all j : $w_j = \sum_{k=1}^{\infty} w_k p_{kj}^{(m)}$

Fundamental limit theorem (3)

- Theorem: an **irreducible aperiodic** Markov chain with transition matrix \mathbf{P} has a **stationary** distribution \mathbf{w} satisfying $w_j > 0$, $\sum_j w_j = 1$, and $\mathbf{w}^T = \mathbf{w}^T \mathbf{P}$ if, and only if, all its states are **positive recurrent**, and this stationary distribution is **unique** and identical to the limiting distribution $w_j = \lim_{n \rightarrow \infty} p_{ij}^{(n)}$ independent of initial state s_i
- Proof (continued):

- have shown $w_j = \sum_{k=1}^{\infty} w_k p_{kj}^{(m)}$

- for $m = 1$, we see the limiting vector \mathbf{w} is **stationary**!!

- next, from $\sum_{j=1}^{\infty} p_{ij}^{(n)} = 1$ then use Fatou:

$$1 = \lim_{n \rightarrow \infty} \sum_{j=1}^{\infty} p_{ij}^{(n)} \geq \sum_{j=1}^{\infty} \lim_{n \rightarrow \infty} p_{ij}^{(n)} = \sum_{j=1}^{\infty} w_j$$

- given $\sum_j w_j \leq 1$ then consider the limit $m \rightarrow \infty$ of

$$w_j = \lim_{m \rightarrow \infty} \sum_{k=1}^{\infty} w_k p_{kj}^{(m)}$$

- since $0 \leq p_{kj}^{(m)} \leq 1$ then $|w_k p_{kj}^{(m)}| \leq w_k$ and $\sum_{k=1}^{\infty} w_k < \infty$ so the dominated convergence theorem can be applied

$$w_j = \lim_{m \rightarrow \infty} \sum_{k=1}^{\infty} w_k p_{kj}^{(m)} = \sum_{k=1}^{\infty} w_k \lim_{m \rightarrow \infty} p_{kj}^{(m)} = \left(\sum_{k=1}^{\infty} w_k \right) w_j$$

- can at last conclude $\sum_{j=1}^{\infty} w_j = 1$

Fundamental limit theorem (4)

- Theorem: an **irreducible aperiodic** Markov chain with transition matrix \mathbf{P} has a **stationary** distribution \mathbf{w} satisfying $w_j > 0$, $\sum_j w_j = 1$, and $\mathbf{w}^T = \mathbf{w}^T \mathbf{P}$ if, and only if, all its states are **positive recurrent**, and this stationary distribution is **unique** and identical to the limiting distribution $w_j = \lim_{n \rightarrow \infty} p_{ij}^{(n)}$ independent of initial state s_i
- Proof (continued):
 - only uniqueness of stationary state to show
 - if another stationary vector \mathbf{v} existed, it would have to satisfy $v_j > 0$, $\sum_{j=1}^{\infty} v_j = 1$, and $v_j = \sum_{i=1}^{\infty} v_i p_{ij}^{(n)}$
 - conditions for dominated convergence theorem again apply, so taking $n \rightarrow \infty$ limit gives
$$v_j = \lim_{n \rightarrow \infty} \sum_{i=1}^{\infty} v_i p_{ij}^{(n)} = \sum_{i=1}^{\infty} v_i \lim_{n \rightarrow \infty} p_{ij}^{(n)} = \left(\sum_{i=1}^{\infty} v_i \right) w_j = w_j$$
 - since $\mathbf{v} = \mathbf{w}$, then \mathbf{w} is unique

An example

- consider the following transition matrix $\mathbf{P} = \begin{bmatrix} \frac{3}{4} & \frac{1}{4} & 0 \\ 0 & \frac{2}{3} & \frac{1}{3} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{bmatrix}$
- \mathbf{P}^2 all positive entries, so chain is irreducible
- eigenvalues are $1, \frac{1}{2}, \frac{5}{12}$
- right and left eigenvectors (unnormalized) are

$$\text{right: } \begin{matrix} 1 & \frac{1}{2} & \frac{5}{12} \\ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} & \begin{bmatrix} 2 \\ -2 \\ 1 \end{bmatrix} & \begin{bmatrix} 3 \\ -4 \\ 3 \end{bmatrix} \end{matrix} \quad \text{left: } \begin{matrix} 1 & \frac{1}{2} & \frac{5}{12} \\ \begin{bmatrix} 2 \\ 3 \\ 2 \end{bmatrix} & \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} & \begin{bmatrix} -3 \\ -1 \\ 4 \end{bmatrix} \end{matrix}$$

- left fixed-point probability vector

$$\mathbf{w} = \frac{1}{7} \begin{bmatrix} 2 \\ 3 \\ 2 \end{bmatrix} \quad \lim_{n \rightarrow \infty} \mathbf{P}^n = \mathbf{W} = \frac{1}{7} \begin{bmatrix} 2 & 3 & 2 \\ 2 & 3 & 2 \\ 2 & 3 & 2 \end{bmatrix}$$

Summary of results

- positive recurrent chain guarantees existence of at least one invariant probability vector
- irreducibility guarantees uniqueness of invariant probability vector
- aperiodicity guarantees limit distribution coincides with invariant distribution

Equilibrium in Markov chains

- suppose a Markov chain is started with probability vector given by \mathbf{w} , the left fixed-point vector of the transition matrix \mathbf{P}
- this means the probability of starting in state s_i is w_i
- the probability of being in state s_j after n steps is $(\mathbf{w}^T \mathbf{P}^n)_j$, but $\mathbf{w}^T \mathbf{P}^n = \mathbf{w}^T$, so this probability is w_j
- thus, the probability vector is always the same, that is, it is **stationary** or **invariant**
- when this occurs, the Markov chain is said to be in **equilibrium**
- recall that an ergodic Markov chain which starts in **any** probability vector \mathbf{y} eventually tends to equilibrium
- the process of bringing the chain into equilibrium from a random starting probability vector is known as **thermalization**

Reversibility in Markov chains

- an ergodic Markov chain is **reversible** if the probability of going from state s_i to s_j is the same as that for going from state s_j to s_i once the chain is in **equilibrium**
- the probability that a transition from s_i to s_j occurs is the probability w_i of finding the chain in state s_i in equilibrium times the transition probability p_{ij}
- reversibility occurs when $w_i p_{ij} = w_j p_{ji}$
- the above condition is often referred to as **detailed balance**
- note that detailed balance guarantees the fixed-point condition: since $\sum_j p_{ij} = 1$ then

$$\sum_j w_j p_{ji} = \sum_j w_i p_{ij} = w_i$$

Law of large numbers for Markov chains

- consider an R -state ergodic Markov chain which starts in state s_i
- define $X_j^{(m)} = \begin{cases} 1 & \text{if chain in state } s_j \text{ after } m \text{ steps} \\ 0 & \text{otherwise} \end{cases}$
- define $N_j^{(n)}$ as number of times chain in state s_j in first n steps

$$N_j^{(n)} = X_j^{(1)} + X_j^{(2)} + \dots + X_j^{(n)}$$

- often called **occupation times**
- expected value $E(X_j^{(m)}) = p_{ij}^{(m)}$ so

$$E(N_j^{(n)}) = \sum_{h=1}^n p_{ij}^{(h)}$$

- it can be shown that

$$\lim_{n \rightarrow \infty} E(N_j^{(n)})/n = w_j$$

- can show **law of large numbers** for ergodic Markov chain:

$$P(|N_j^{(n)}/n - w_j| > \varepsilon) \rightarrow 0 \text{ as } n \rightarrow \infty$$

Central limit and ergodic theorem for Markov chains

- can show a central limit holds

$$\lim_{n \rightarrow \infty} P \left(\frac{a\sigma_j}{\sqrt{n}} < \left(\frac{N_j^{(n)}}{n} - w_j \right) < \frac{b\sigma_j}{\sqrt{n}} \right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx$$

where σ_j depends on w_j

- distributions of random variables $N_j^{(n)}$ tend to normal distributions
- let X_1, X_2, \dots, X_n be the actual outcomes that make up an ergodic R -state Markov chain
- from the definition of $X_j^{(n)}$, it follows that $\sum_{j=1}^R X_j^{(n)} = 1$ so

$$\frac{1}{n} \sum_{h=0}^{n-1} f(X_h) = \frac{1}{n} \sum_{h=0}^{n-1} \sum_{j=1}^R X_j^{(h)} f(s_j) = \sum_{j=1}^R N_j^{(n)} f(s_j) \rightarrow \sum_{j=1}^R w_j f(s_j)$$

- Markov-chain “time”-average approaches required ensemble average!!
- already knew this for stationary stochastic processes

Monte Carlo integration

- Monte Carlo integration using Markov chain in equilibrium:

$$\int_V p(\vec{x}) f(\vec{x}) d^D x \approx \langle f \rangle \pm \sqrt{\frac{R_0(f) + 2 \sum_{h \geq 1} R_h(f)}{N}}$$

$$\langle f \rangle \equiv \frac{1}{N} \sum_{i=1}^N f(\vec{x}_i), \quad R_h(f) \equiv \frac{1}{N-h} \sum_{i=1}^{N-h} (f(\vec{x}_i) - \langle f \rangle) (f(\vec{x}_{i+h}) - \langle f \rangle)$$

- each point in D -dim. volume V is a state of a Markov chain
- N points $\vec{x}_1, \dots, \vec{x}_N$ are elements of an irreducible aperiodic Markov chain with positive recurrent states and **stationary** or limiting probability dist. $p(\vec{x})$ throughout D -dimensional volume V
- Markov chain should be in equilibrium
- normalization condition $\int_V p(\vec{x}) d^D x = 1$
- absolutely summable autocovariance $\sum_{h=0}^{\infty} |R_h(f)| < \infty$

Autocorrelations

- configurations generated by Markov process depend on previous elements in the chain
- this dependence known as **autocorrelation**
- this autocorrelation can actually be measured!
- for any observable (integrand) O_i , autocorrelation $\rho(\tau)$ defined by

$$\frac{\langle O_i O_{i+\tau} \rangle - \langle O_i \rangle^2}{\langle O_i^2 \rangle - \langle O_i \rangle^2}$$

- highly correlated \rightarrow value near 1
 - independent \rightarrow value near 0
- decreasing autocorrelations decreases Monte Carlo error
- dependence decreases as distance between elements in chain increases
 - do not use every element in chain for "measurements"
 - skip some number of elements between measurements

Constructing the transition probability

- generally know probability density $\pi(\phi)$ we need to sample
- for our path integrals, we need to generate paths with probability distribution

$$\pi(\phi) = \frac{e^{-S[\phi]/\hbar}}{\int_a^b \mathcal{D}\phi' e^{-S[\phi']/\hbar}}$$

- in imaginary time formalism, path integral weight is real and positive \rightarrow probability interpretation for Monte Carlo
- how do we construct the Markov transition matrix $P(\tilde{\phi} \leftarrow \phi)$?
 - change to quantum mechanical notation of putting earlier states on right, later states on left
- simplest answer to this question is

the **Metropolis-Hastings** method

- useful for local updating so changes to action are small
- probability normalization never enters in the calculation!

The Metropolis-Hastings algorithm

- this method uses an auxiliary **proposal** density $R(\tilde{\phi} \leftarrow \phi)$ which
 - must be normalized
 - can be evaluated for all $\phi, \tilde{\phi}$
 - can be easily sampled
 - no relationship to the fixed-point probability density $\pi(\phi)$ needed
- given this proposal density, the Metropolis-Hastings method updates $\phi \rightarrow \tilde{\phi}$ as follows:
 - 1 use $R(\tilde{\phi} \leftarrow \phi)$ to propose new value $\tilde{\phi}$ from current value ϕ
 - 2 accept the new value with probability

$$P_{\text{acc}}(\tilde{\phi} \leftarrow \phi) = \min \left(1, \frac{R(\phi \leftarrow \tilde{\phi})\pi(\tilde{\phi})}{R(\tilde{\phi} \leftarrow \phi)\pi(\phi)} \right)$$

- 3 if rejected, the original value ϕ is retained
- if proposal density satisfies reversibility $R(\tilde{\phi} \leftarrow \phi) = R(\phi \leftarrow \tilde{\phi})$, then acceptance probability reduces to $\min(1, \pi(\tilde{\phi})/\pi(\phi))$
 - original Metropolis method

Detailed balance in Metropolis-Hastings

- Metropolis-Hastings satisfies detailed balance
- proof:

- (normalized) transition probability density is

$$W(\tilde{\phi} \leftarrow \phi) = P_{\text{acc}}(\tilde{\phi} \leftarrow \phi)R(\tilde{\phi} \leftarrow \phi) + \delta(\tilde{\phi} - \phi) \left(1 - \int \mathcal{D}\bar{\phi} P_{\text{acc}}(\bar{\phi} \leftarrow \phi)R(\bar{\phi} \leftarrow \phi) \right)$$

- define

$$\begin{aligned} A(\tilde{\phi} \leftarrow \phi) &\equiv P_{\text{acc}}(\tilde{\phi} \leftarrow \phi)R(\tilde{\phi} \leftarrow \phi)\pi(\phi) \\ &= \min \left(1, \frac{R(\phi \leftarrow \tilde{\phi})\pi(\tilde{\phi})}{R(\tilde{\phi} \leftarrow \phi)\pi(\phi)} \right) R(\tilde{\phi} \leftarrow \phi)\pi(\phi) \\ &= \min \left(R(\tilde{\phi} \leftarrow \phi)\pi(\phi), R(\phi \leftarrow \tilde{\phi})\pi(\tilde{\phi}) \right) \end{aligned}$$

where last line follows from $R(\tilde{\phi} \leftarrow \phi)\pi(\phi) \geq 0$

- symmetric: $A(\tilde{\phi} \leftarrow \phi) = A(\phi \leftarrow \tilde{\phi})$.

Detailed balance in Metropolis-Hastings (continued)

- so we have

$$\begin{aligned}W(\tilde{\phi} \leftarrow \phi)\pi(\phi) &= P_{\text{acc}}(\tilde{\phi} \leftarrow \phi)R(\tilde{\phi} \leftarrow \phi)\pi(\phi) \\ &+ \delta(\tilde{\phi} - \phi) \left(1 - \int \mathcal{D}\bar{\phi} P_{\text{acc}}(\bar{\phi} \leftarrow \phi)R(\bar{\phi} \leftarrow \phi) \right) \pi(\phi) \\ &= A(\tilde{\phi} \leftarrow \phi) + \delta(\tilde{\phi} - \phi) \left(\pi(\phi) - \int \mathcal{D}\bar{\phi} A(\bar{\phi} \leftarrow \phi) \right) \\ &= A(\tilde{\phi} \leftarrow \phi) + \delta(\tilde{\phi} - \phi) K(\phi)\end{aligned}$$

where

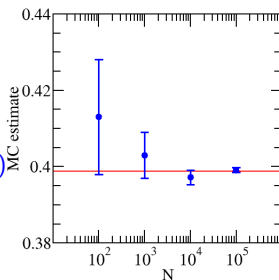
$$K(\phi) = \pi(\phi) - \int \mathcal{D}\bar{\phi} A(\bar{\phi} \leftarrow \phi)$$

- given symmetry of A and Dirac δ -function, then detailed balance holds

$$W(\tilde{\phi} \leftarrow \phi)\pi(\phi) = W(\phi \leftarrow \tilde{\phi})\pi(\tilde{\phi})$$

A one dimensional example

- does this really work?
- let $g(x) = \cos(\sqrt{1+x^2})$ and $h(x) = e^{-x^2}/(x^2+2)$
- $g(x)$ changes sign, $h(x) \geq 0$
- consider ratio of integrals $I = \frac{\int_{-\infty}^{\infty} g(x)h(x)dx}{\int_{-\infty}^{\infty} h(x)dx} = 0.3987452$
- sampling density $\pi(x) = Z^{-1}h(x)$ where $Z = \int_{-\infty}^{\infty} h(x)dx$
- algorithm:
 - choose δ uniform probability for $-\Delta \leq \delta \leq \Delta$
 - propose $\tilde{x} = x + \delta$
 - acceptance probability $\min(1, \pi(\tilde{x})/\pi(x)) = \min(1, h(\tilde{x})/h(x))$
- $\Delta = 1.5$ for acceptance $\sim 50\%$
- never needed Z



Part III

Monte Carlo study of the simple harmonic oscillator

Discretization of SHO action

- action of harmonic oscillator (imaginary time formalism)

$$S[x(\tau)] = \int_{\tau_a}^{\tau_b} d\tau \left(\frac{1}{2}m\dot{x}^2 + \frac{1}{2}m\omega^2 x^2 \right)$$

- discretize time $N\varepsilon = \tau_b - \tau_a$ for Monte Carlo evaluation

$$\frac{S}{\hbar} = \frac{m\varepsilon}{2\hbar} \sum_{j=0}^{N-1} \left[\left(\frac{x_{j+1} - x_j}{\varepsilon} \right)^2 + \omega^2 \left(\frac{x_{j+1} + x_j}{2} \right)^2 \right]$$

- choose ε so discretization errors sufficiently small
- introduce dimensionless parameters

$$x_k = d_k \sqrt{\frac{\varepsilon \hbar}{m}} \quad \kappa = \frac{1}{4} \varepsilon^2 \omega^2$$
$$\frac{S}{\hbar} = \frac{1}{2} \sum_{j=0}^{N-1} \left[(d_{j+1} - d_j)^2 + \kappa (d_{j+1} + d_j)^2 \right]$$

Discretization of action (continued)

- a few more manipulations produce

$$\frac{S}{\hbar} = \frac{1}{2}(1 + \kappa)(d_0^2 + d_N^2) + (1 + \kappa) \left[\sum_{j=1}^{N-1} d_j^2 \right] - (1 - \kappa) \left[\sum_{j=0}^{N-1} d_j d_{j+1} \right]$$

- first constant irrelevant (set to zero), then one last rescaling

$$u_j = d_j \sqrt{1 + \kappa} \quad g = \frac{1 - \kappa}{1 + \kappa} \quad d_0 = d_N = 0$$

- final result for action

$$\frac{S}{\hbar} = \left[\sum_{j=1}^{N-1} u_j^2 \right] - g \left[\sum_{j=0}^{N-1} u_j u_{j+1} \right]$$

Metropolis updating of path

- to update location (at a single time)
 - propose random shift $-\Delta \leq \delta \leq \Delta$ with uniform probability
 - calculate change to the action
$$\delta S/\hbar = \delta (\delta + 2u_j - g(u_{j-1} + u_{j+1}))$$
 - accept $u_j^{\text{new}} = u_j + \delta$ with probability $\min(1, e^{-\delta S/\hbar})$
 - rule of thumb: fix Δ for about 50% acceptance rate
 - lower rate = wasting too much time with rejections
 - higher rate = moving through phase space too slowly
- repeat for each u_j for $j = 1, \dots, N - 1$ (this is called one **sweep**)
- repeat for certain number of sweeps
 - until autocorrelations sufficiently small

Actual C++ code

- here is actual C++ code which does the updating

```
void markov::update()
{
    double shift,deltaS;
    for (int i=1;i<=Nsweeps;i++)
        for (int t=1;t<Ntimesteps;t++){
            // propose shift in location[t]
            shift=2.0*max_shift_per_instance*(rng.generate()-0.5);

            // compute change in action
            deltaS=shift*(shift+2.0*locations[t]
                -hop_param*(locations[t-1]+locations[t+1]));

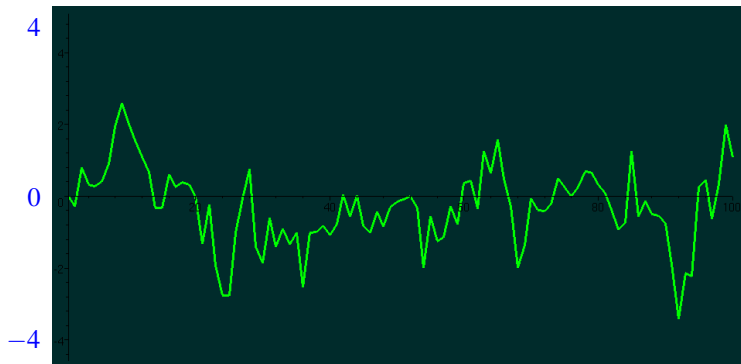
            // Metropolis accept or reject
            if (deltaS<0.0) accept=1;
            else accept=(rng.generate()<=exp(-deltaS));
            if (accept) locations[t]+=shift;
        }
}
```


Simulation guidelines

- to start Markov chain
 - choose a random path (hot start)
 - or choose $u_j = 0$ for all j (cold start)
 - update N_{therm} sweeps until fixed point of chain achieved (thermalization) → check some simple observable
- once thermalized, begin “measurements”
- must choose
 - ε so discretization errors sufficiently small
 - Δ for adequate acceptance rate
 - N_{sweeps} for sufficiently small autocorrelations
 - N_{meas} for desired precision of results

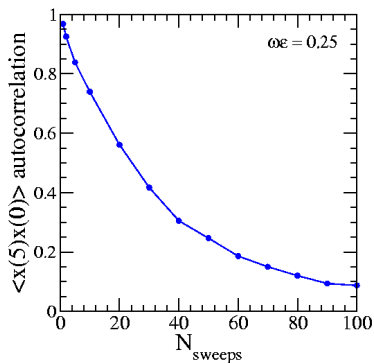
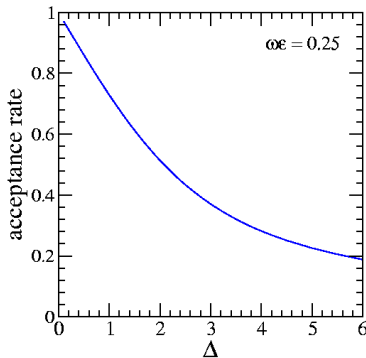
Path animation

- animation of first 100 time slices of u_j path



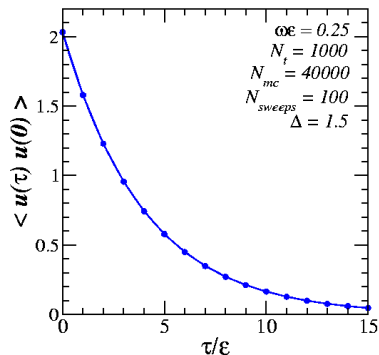
Acceptance rate and autocorrelations

- choose Δ so acceptance rate near 0.5
- choose N_{sweeps} so autocorrelations near 0.1



Correlation function

- comparison of final Monte Carlo estimates with exact results



- exact result shown as curve
- Monte Carlo estimates shown by circles (statistical uncertainties too small to see)

Part IV

Monte Carlo calculations in real scalar field theory in 2+1 dimensions

Action in continuous space-time

- action in continuous Euclidean D -dimensional space-time (imaginary time formalism) given by

$$S = \int d^D x \left(\frac{1}{2} \partial_\mu \varphi(x) \partial_\mu \varphi(x) + \frac{1}{2} m^2 \varphi(x)^2 + \frac{g}{4!} \varphi(x)^4 \right).$$

- action must be dimensionless (natural units $\hbar = c = 1$)
- m has units of a derivative ∂_μ , that is, of a mass
- units of field $[\phi] = [m]^{\frac{1}{2}D-1}$
- coupling g has units $[g] = [m]^{4-D}$
 - coupling dimensionless in 4 space-time dimensions
 - has units of mass in 3 space-time dimensions so g/m dimensionless

Quantization

- quantization using path integrals
- generalize notion of “path”: a path here is a field configuration
- path integral is now integrations over all field configurations
- for real scalar field, integral $-\infty \leq \phi(x) \leq \infty$ at every space-time point x
- time-ordered two-point function given by

$$\langle T\phi(x_1)\phi(x_2) \rangle = \frac{\int \mathcal{D}\phi \phi(x_1)\phi(x_2) \exp(-S[\phi])}{\int \mathcal{D}\phi \exp(-S[\phi])}.$$

- generalizes to n -point functions: time-ordered product of n fields

Discretization of action

- Monte Carlo study requires action on a space-time lattice
- use anisotropic cubic lattice with temporal lattice spacing a_t and spatial lattice spacing a_s
- use simplest finite difference for the field derivatives
- action is given by

$$\begin{aligned} S &= a_s^{D-1} a_t \sum_x \left(\sum_{\mu} \frac{(\varphi(x+a_{\mu}\hat{\mu})-\varphi(x))^2}{2a_{\mu}^2} + \frac{1}{2}m^2\varphi(x)^2 + \frac{g}{4!}\varphi(x)^4 \right) \\ &= a_s^{D-1} a_t \sum_x \left(-\sum_{\mu} \frac{\varphi(x+a_{\mu}\hat{\mu})\varphi(x)}{a_{\mu}^2} + \frac{1}{2} \left(m^2 + \sum_{\nu} \frac{2}{a_{\nu}^2} \right) \varphi(x)^2 + \frac{g}{4!}\varphi(x)^4 \right) \end{aligned}$$

- redefine the field: $\sqrt{a_s^{D-3} a_t} \varphi(x) = \sqrt{2\kappa_s} \phi(x)$
where κ_s is dimensionless number, new field $\phi(x)$ is dimensionless

Action on lattice

- a few more dimensionless parameters:

$$a_s/a_t = \zeta, \quad \lambda = \frac{g\zeta\kappa_s^2}{6a_s^{D-4}},$$

$$\kappa_s(a_s^2 m^2 + 2\zeta^2 + 2D - 2) = 1 - 2\lambda, \quad \kappa = \zeta\kappa_s$$

- final form for lattice action

$$S = \sum_x \left(-\frac{2\kappa}{\zeta} \sum_{j=1}^{D-1} \phi(x)\phi(x+a_s\hat{j}) - 2\kappa\zeta \phi(x)\phi(x+a_t\hat{t}) \right. \\ \left. + (1 - 2\lambda)\phi(x)^2 + \lambda\phi(x)^4 \right)$$

- hopping parameter κ essentially sets mass parameter, λ is interaction coupling

Exact results in free field theory

- the free field theory $\lambda = 0$ is exactly solvable
- path integrals are multivariate gaussians
- free action can be written in form

$$S[\phi] = \frac{1}{2} \sum_{xy} \phi(x) M(x, y) \phi(y)$$

- for N lattice sites, M is real and symmetric $N \times N$ matrix having positive eigenvalues
- this matrix given by

$$M(x, y) = -\frac{2\kappa}{\zeta} \sum_{j=1}^{D-1} (\delta(y, x+a_s \hat{j}) + \delta(x, y+a_s \hat{j})) \\ - 2\kappa\zeta (\delta(y, x+a_t \hat{t}) + \delta(x, y+a_t \hat{t})) + 2\delta(x, y)$$

Gaussian integrals in free theory

- N -dimensional multivariate Gaussian integral of form

$$\begin{aligned} & \prod_{i=1}^N \left(\int_{-\infty}^{\infty} d\phi_i \right) \exp\left(-\frac{1}{2}\phi_j M_{jk} \phi_k + J_n \phi_n\right) \\ &= \left(\det \left(\frac{M}{2\pi} \right) \right)^{-1/2} \exp\left(\frac{1}{2} J_j M_{jk}^{-1} J_k\right) \end{aligned}$$

- J -trick: use derivatives wrt to J_k , followed by $J_k \rightarrow 0$ to evaluate all integrals involving any number of products of the fields

$$\begin{aligned} & \prod_{i=1}^N \left(\int_{-\infty}^{\infty} d\phi_i \right) \phi_{m_1} \phi_{m_2} \dots \phi_{m_r} \exp\left(-\frac{1}{2}\phi_j M_{jk} \phi_k\right) \\ &= \frac{\delta}{\delta J_{m_1}} \dots \frac{\delta}{\delta J_{m_r}} \prod_{i=1}^N \left(\int_{-\infty}^{\infty} d\phi_i \right) \exp\left(-\frac{1}{2}\phi_j M_{jk} \phi_k + J_n \phi_n\right) \end{aligned}$$

- does Wick contractions automatically!

Two-point function

- two-point function given by $\langle T\phi(x_1)\phi(x_2) \rangle = M^{-1}(x_1, x_2)$
- invert M by method of Green functions and use Fourier series
- for $L_x \times L_y \times L_t$ lattice, result is

$$M^{-1}(x, y) = \frac{\zeta}{2\kappa L_x L_y L_t} \sum_{k_\mu} \frac{\cos(k \cdot (x - y))}{(a_s^2 m^2 + 4 \sum_{j=1}^2 \sin^2(\frac{1}{2} k_j) + 4\zeta^2 \sin^2(\frac{1}{2} k_t))}$$

where $k_\mu = 2\pi n_\mu / L_\mu$ for $n_\mu = 0, 1, 2, \dots, L_\mu - 1$

- pole gives energy $a_t E_p$ of single particle of momentum $a_s p$

$$a_t E_p = 2 \sinh^{-1} \left(\frac{1}{2\zeta} \sqrt{a_s^2 m^2 + 4 \sin^2(\frac{1}{2} a_s p_x) + 4 \sin^2(\frac{1}{2} a_s p_y)} \right)$$

- for small a_t, a_s this becomes $E_p = \sqrt{m^2 + p_x^2 + p_y^2}$
- spectrum is sum of free particle energies

Single-site Monte Carlo updating

- Metropolis-Hastings method needs acceptable acceptance rate
- changing all field values at once generally leads to large changes in action \rightarrow near zero acceptance rate
- reasonable acceptance rate achieved by updating field at a **single** lattice site at any given time
- ergodicity ensured by **sweeping** through lattice, updating each and every site one at a time
- in battle against autocorrelations, expect
 - small wavelength modes updated well
 - long wavelength modes updated not so well

δS for single-site update

- recall action is

$$S = \sum_x \left(-\frac{2\kappa}{\zeta} \sum_{j=1}^{D-1} \phi(x) \phi(x+a_s \hat{j}) - 2\kappa\zeta \phi(x) \phi(x+a_t \hat{t}) \right. \\ \left. + (1 - 2\lambda) \phi(x)^2 + \lambda \phi(x)^4 \right)$$

- for $\tilde{\phi} \leftarrow \phi$, change in action is $\delta S = S[\tilde{\phi}] - S[\phi]$
- define neighborhood

$$N(x) = -\frac{2\kappa}{\zeta} \sum_{j=1}^{D-1} \left(\phi(x+a_s \hat{j}) + \phi(x-a_s \hat{j}) \right) - 2\kappa\zeta \left(\phi(x+a_t \hat{t}) + \phi(x-a_t \hat{t}) \right)$$

- if field at one site x changed $\phi(x) \rightarrow \phi(x) + \Delta$, then

$$\delta S = \Delta \left(N(x) + (\Delta + 2\phi(x)) \left(1 + \lambda \left((\Delta + 2\phi(x)) \Delta + 2(\phi(x)^2 - 1) \right) \right) \right)$$

δS for single-site update (continued)

- change in action can also be written

$$\delta S = \Delta (a_0 + a_1 \Delta + a_2 \Delta^2 + a_3 \Delta^3),$$

$$a_0 = N(x) + 2\phi(x)(1 + 2\lambda(\phi(x)^2 - 1))$$

$$a_1 = 1 + 2\lambda(3\phi(x)^2 - 1)$$

$$a_2 = 4\lambda\phi(x)$$

$$a_3 = \lambda$$

Metropolis sweeps

- single-site updates involve a single continuous real variable ϕ
- use simplest proposal density

$$R(\tilde{\phi} \leftarrow \phi) = \begin{cases} \frac{1}{\Delta_0} & -\frac{1}{2}\Delta_0 \leq (\tilde{\phi} - \phi) \leq \frac{1}{2}\Delta_0 \\ 0 & |\tilde{\phi} - \phi| > \frac{1}{2}\Delta_0 \end{cases}$$

- width Δ_0 chosen for acceptance probability around 50%
- proposed new value accepted with probability $\min(1, \exp(-\delta S))$
- if rejected, keep current field value
- sweeping through lattice ensures ergodicity
- in sweeping through the lattice in predetermined order, detailed balance no longer holds
 - not a problem since the fixed-point stability condition still holds
 - detailed balance maintained by updating sites in random order

Battling autocorrelations

- when the single particle mass $a_t m_{\text{gap}}$ is small, the **coherence length** $\xi = 1/(a_t m_{\text{gap}})$ becomes large
- $\xi \rightarrow \infty$ signals continuum limit
- $\xi \rightarrow \infty$ occurs near critical point (2nd order phase transition)
- we will see that autocorrelations with Metropolis updating become long ranged as ξ becomes large
 - known as **critical slowing down**
- autocorrelations problematic even for $\xi \approx 5$ with Metropolis
- need help to better update long wavelength modes

Microcanonical updating

- long wavelength modes are associated with lower frequencies, lower energies
- in other words, long-wavelength modes associated with very small changes to the action
- possible way to improve autocorrelations:
 - make large but action preserving $\delta S = 0$ changes to field at one site
- call this a microcanonical update
 - often referred to as overrelaxation
- local updating is so easy, don't want to give up on it yet!
- must still update in such a way to satisfy detailed balance
- not ergodic, so microcanonical sweeps must be used in combination with ergodic scheme, such as Metropolis sweeps

Microcanonical updating (2)

- we know Metropolis-Hasting method satisfies detailed balance
- choose proposal density strongly peaked about action-preserving value of field, then carefully take δ -function limit
- revisit Metropolis-Hastings with sharply-peaked Breit-Wigner proposal probability density

$$R_f(\tilde{\phi} \leftarrow \phi) = \frac{1}{\pi} \frac{\varepsilon}{(\tilde{\phi} - f(\phi))^2 + \varepsilon^2}$$

where ε is a constant and $f(\phi)$ is well-behaved, single-valued, invertible function

- acceptance probability

$$P_{\text{acc}}(\tilde{\phi} \leftarrow \phi) = \min\left(1, \frac{R_f(\phi \leftarrow \tilde{\phi})\pi(\tilde{\phi})}{R_f(\tilde{\phi} \leftarrow \phi)\pi(\phi)}\right) = \min\left(1, \frac{((\tilde{\phi} - f(\phi))^2 + \varepsilon^2)\pi(\tilde{\phi})}{((\phi - f(\tilde{\phi}))^2 + \varepsilon^2)\pi(\phi)}\right)$$

Microcanonical updating (3)

- carefully take $\varepsilon \rightarrow 0$ limit: $R_f(\tilde{\phi} \leftarrow \phi) \rightarrow \delta(\tilde{\phi} - f(\phi))$
- determining acceptance probability is tricky
- probability of proposing a value between $f(\phi) - \sqrt{\varepsilon} \leq \tilde{\phi} \leq f(\phi) + \sqrt{\varepsilon}$ is

$$\int_{f(\phi) - \sqrt{\varepsilon}}^{f(\phi) + \sqrt{\varepsilon}} d\tilde{\phi} R_f(\tilde{\phi} \leftarrow \phi) = \frac{2}{\pi} \tan^{-1} \left(\frac{1}{\sqrt{\varepsilon}} \right)$$

which does tends to unity as $\varepsilon \rightarrow 0$

- if $f(\phi)$ more than $\sqrt{\varepsilon}$ away from ϕ , probability transition is actually made is

$$\begin{aligned} \int_{f(\phi) - \sqrt{\varepsilon}}^{f(\phi) + \sqrt{\varepsilon}} d\tilde{\phi} W_f(\tilde{\phi} \leftarrow \phi) &= \int_{f(\phi) - \sqrt{\varepsilon}}^{f(\phi) + \sqrt{\varepsilon}} d\tilde{\phi} P_{\text{acc}}(\tilde{\phi} \leftarrow \phi) R_f(\tilde{\phi} \leftarrow \phi) \\ &= \min \left(\frac{2}{\pi} \tan^{-1} \left(\frac{1}{\sqrt{\varepsilon}} \right), \frac{1}{\pi} \int_{f(\phi) - \sqrt{\varepsilon}}^{f(\phi) + \sqrt{\varepsilon}} d\tilde{\phi} \frac{\varepsilon \pi(\tilde{\phi})}{((\phi - f(\tilde{\phi}))^2 + \varepsilon^2) \pi(\phi)} \right) \end{aligned}$$

Microcanonical updating (4)

- write $\tilde{\phi} = f(\phi) + y$, then remaining integral becomes

$$\frac{1}{\pi} \int_{-\sqrt{\varepsilon}}^{\sqrt{\varepsilon}} dy \frac{\varepsilon \pi(f(\phi) + y)}{\left((\phi - f(f(\phi) + y))^2 + \varepsilon^2 \right) \pi(\phi)}$$

- if $f(f(\phi)) \neq \phi$, can show this integral goes to zero as $\varepsilon \rightarrow 0$
- for **self-inverse** function $f(f(\phi)) = \phi$, expansion about $y = 0$ must be carefully done, integral has form

$$\frac{\varepsilon}{\pi} \int_{-\sqrt{\varepsilon}}^{\sqrt{\varepsilon}} dy \frac{(a_0 + a_1 y + a_2 y^2 + \dots)}{(\varepsilon^2 + b_2 y^2 + b_3 y^3 + b_4 y^4 \dots)}$$

- must retain $b_2 y^2$ in denominator, expand rest about $y = 0$:

$$\frac{\varepsilon}{\pi} \int_{-\sqrt{\varepsilon}}^{\sqrt{\varepsilon}} dy \frac{a_0}{(\varepsilon^2 + b_2 y^2)} \left\{ 1 + \frac{a_1}{a_0} y + \frac{a_2}{a_0} y^2 + \left(\frac{a_3}{a_0} - \frac{b_3}{\varepsilon^2} \right) y^3 \dots \right\}$$

- for $b_2 > 0$, result of integration is

$$\frac{2a_0}{\pi\sqrt{b_2}} \tan^{-1} \left(\sqrt{\frac{b_2}{\varepsilon}} \right) \left\{ 1 + d_1 \sqrt{\varepsilon} + d_2 \varepsilon + d_3 \varepsilon^{3/2} + \dots \right\}$$

Microcanonical updating (5)

- acceptance probability in limit $\varepsilon \rightarrow 0$ given by

$$P_{\text{acc}} = \min \left(1, \frac{a_0}{\sqrt{b_2}} \right)$$

- here $a_0 = \pi(f(\phi))/\pi(\phi)$ and $b_2 = (f'(f(\phi)))^2$
- differentiate both sides of $f(f(\phi)) = \phi$ with respect to ϕ , so for self-inverse function

$$1 = \frac{d}{d\phi} \left(f(f(\phi)) \right) = f'(f(\phi)) f'(\phi)$$

$$\frac{1}{(f'(f(\phi)))^2} = \left| \frac{f'(\phi)}{f'(f(\phi))} \right| \quad (\text{self-inverse function})$$

- take limit $\varepsilon \rightarrow 0$ acceptance probability goes to

$$P_{\text{acc}}(\tilde{\phi} \leftarrow \phi) = \min \left(1, \frac{\sqrt{|f'(\phi)|} \pi(\tilde{\phi})}{\sqrt{|f'(\tilde{\phi})|} \pi(\phi)} \right)$$

Microcanonical updating (6)

- specialize to **action preserving** function $f(\phi)$
- for infinitesimal change $\phi \rightarrow \phi + \delta\phi$

$$S(\phi + \delta\phi) = S(f(\phi + \delta\phi))$$

- expand both sides

$$\begin{aligned} S(\phi) + S'(\phi)\delta\phi + O(\delta\phi^2) &= S(f(\phi) + f'(\phi)\delta\phi + O(\delta\phi^2)) \\ &= S(f(\phi)) + S'(f(\phi))f'(\phi)\delta\phi + O(\delta\phi^2) \\ &= S(\phi) + S'(f(\phi))f'(\phi)\delta\phi + O(\delta\phi^2). \end{aligned}$$

- solve order by order in $\delta\phi$

$$S'(\phi) = S'(f(\phi))f'(\phi) \rightarrow f'(\phi) = \frac{S'(\phi)}{S'(f(\phi))}, \quad f'(f(\phi)) = \frac{S'(f(\phi))}{S'(\phi)}$$

- proposal and acceptance probability densities are

$$R_f(\tilde{\phi} \leftarrow \phi) = \delta(\tilde{\phi} - f(\phi)), \quad f(f(\phi)) = \phi, \quad S(f(\phi)) = S(\phi),$$

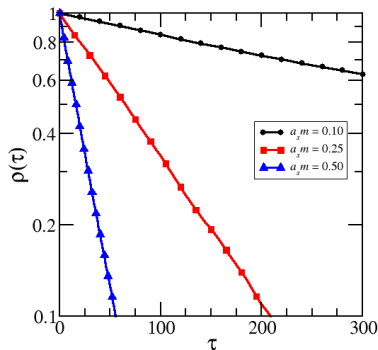
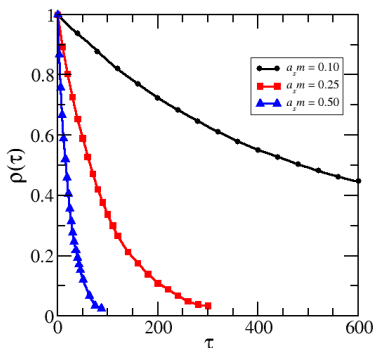
$$P_{\text{acc}}(\tilde{\phi} \leftarrow \phi) = \min\left(1, \left|\frac{S'(\phi)}{S'(\tilde{\phi})}\right|\right), \quad \pi(\phi) = \frac{\exp(-S[\phi])}{\int \mathcal{D}\tilde{\phi} \exp(-S[\tilde{\phi}])}$$

Microcanonical updating (7)

- generalize to multiple self-inverse functions
 - for ϕ^4 at most four field values with same local action
- generalize to probability μ of proposing a change
 - sometimes need $\mu < 1$ to prevent (damped) oscillations in autocorrelation function
- summary of microcanonical updating process:
 - 1 decide to propose new field value with probability μ (skip steps below if no proposal)
 - 2 solve $\delta S(\phi) = 0$, let ϕ_j denote real solutions different from ϕ
 - these are roots of a cubic polynomial
 - 3 randomly choose one of the ϕ_j with equal probability, let $\tilde{\phi}$ denote the chosen value
 - 4 accept with probability
$$P_{\text{acc}}(\tilde{\phi} \leftarrow \phi) = \min \left(1, \left| \frac{S'(\phi)}{S'(\tilde{\phi})} \right| \right)$$
if rejected, original value ϕ retained

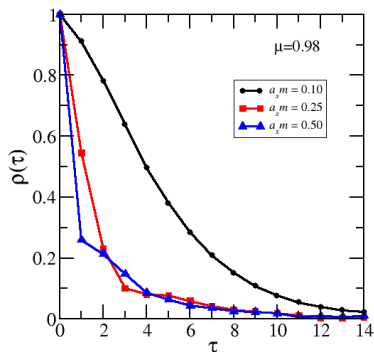
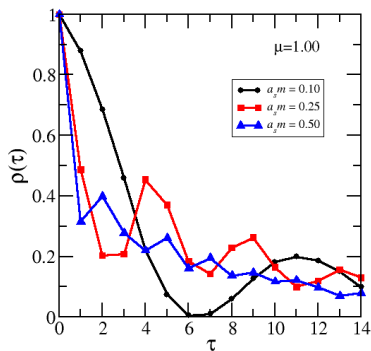
Autocorrelations

- studied autocorrelation function $\rho(\tau)$ of $\langle \Phi(t)\Phi(0) \rangle$ for $t = 1/(2a_s m)$ and $\Phi(t) = \sum_{xy} \phi(x, y, t)$
- τ is number of Metropolis sweeps in plots below
- $a_s m = 0.10, 0.25, 0.50$ for $\lambda = 0$ on 24^3 isotropic lattice
- 2200 sweeps to reduce autocorrelations to 0.1 for $a_s m = 0.10$



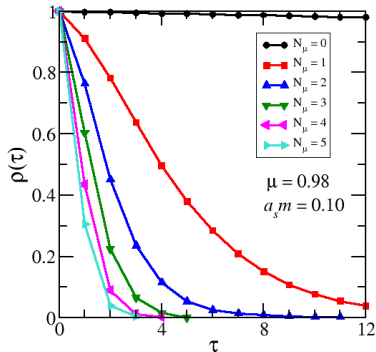
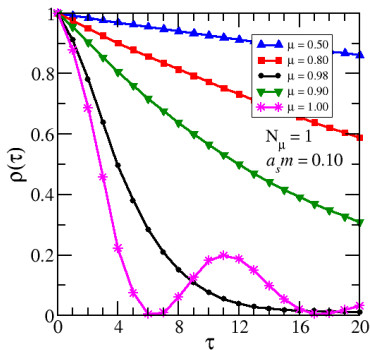
Autocorrelations

- autocorrelations $\rho(\tau)$ of $\langle \Phi(t)\Phi(0) \rangle$ for $t = 1/(2a_s m)$
- τ is number of compound sweeps,
- compound sweep = 1 Metropolis + 1 microcanonical sweep
- μ is probability of proposing change in microcanonical updates
- $a_s m = 0.10, 0.25, 0.50$ for $\lambda = 0$ on 24^3 isotropic lattice
- undesirable oscillations on left removed using $\mu = 0.98$ or updating sites in random order



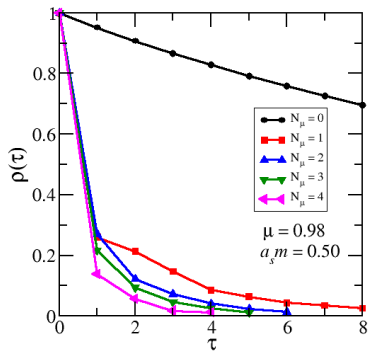
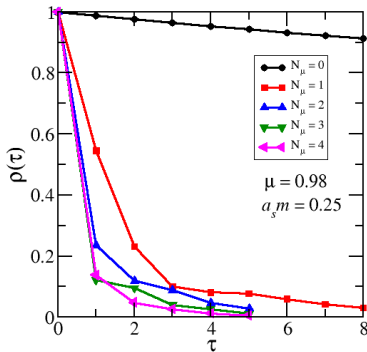
Autocorrelations

- autocorrelations $\rho(\tau)$ of $\langle \Phi(t)\Phi(0) \rangle$ for $t = 1/(2a_s m)$
- τ is number of compound sweeps
- compound sweep = 1 Metropolis + N_μ microcanonical sweeps
- μ is probability of proposing change in microcanonical updates
- $a_s m = 0.10$ for $\lambda = 0$ on 24^3 isotropic lattice
- left-hand plot, $N_\mu = 1$ and μ is varied
- right-hand plot, $\mu = 0.98$ and N_μ is varied



Autocorrelations

- autocorrelations $\rho(\tau)$ of $\langle \Phi(t)\Phi(0) \rangle$ for $t = 1/(2a_s m)$
- τ is number of compound sweeps
- compound sweep = 1 Metropolis + N_μ microcanonical sweeps
- $\mu = 0.98$ probability of proposing change in microcanonical
- $a_s m = 0.25, 0.50$ for $\lambda = 0$ on 24^3 isotropic lattice



Calculating the spectrum

- stationary-state energies extracted from asymptotic decay rates of temporal correlations of the fields
- temporal evolution of field as Heisenberg-picture quantum operator

$$\phi(t) = e^{Ht} \phi(0) e^{-Ht}$$

- under certain general assumptions and ignoring temporal boundary conditions, then for $t \geq 0$

$$\begin{aligned} \langle 0 | \phi(t) \phi(0) | 0 \rangle &= \sum \langle 0 | e^{Ht} \phi(0) e^{-Ht} | n \rangle \langle n | \phi(0) | 0 \rangle, \\ &= \sum^n \left| \langle n | \phi(0) | 0 \rangle \right|^2 e^{-(E_n - E_0)t} = \sum A_n e^{-(E_n - E_0)t}, \end{aligned}$$

- where complete set of (discrete) eigenstates of H^n satisfying $H|n\rangle = E_n|n\rangle$ inserted
- if $\langle 1 | \phi(0) | 0 \rangle \neq 0$, then A_1 and $E_1 - E_0$ can be extracted as t becomes large, assuming $\langle 0 | \phi(0) | 0 \rangle = 0$
- can use any operator $O(t)$ which is a function of the field $\phi(t)$ only on a time slice t

Calculating the spectrum (2)

- extraction of A_1 and $E_1 - E_0$ done using correlated χ^2

$$\chi^2 = \sum_{t'} \left(C(t) - M(t, \alpha) \right) \sigma_{t'}^{-1} \left(C(t') - M(t', \alpha) \right)$$

where $C(t)$ represents Monte Carlo estimates of correlation function with covariance matrix $\sigma_{t'}$ and model function is $M(t, \alpha) = \alpha_1 e^{-\alpha_0 t}$.

- minimize expression with respect to the model parameters α_0, α_1
- uncertainties in the best-fit parameters $\alpha_0 = E_1 - E_0$ and $\alpha_1 = A_1$ are obtained by a **jackknife** or **bootstrap** procedure
- fit must be done for a time range $t_{\min} \leq t \leq t_{\max}$ such that an acceptable fit quality is obtained, that is, $\chi^2/\text{dof} \approx 1$
- sum of two-exponentials as model function can be used to minimize sensitivity to t_{\min}
 - but fit parameters associated with faster-decaying exponential generally *not* good estimates of gap to next energy level and should be discarded

Jackknife resampling

- return to independent trials process X_1, X_2, \dots, X_N
- expected value $E(f(X))$ estimated using $\langle f \rangle = \frac{1}{N} \sum_{k=1}^N f(X_k)$
- sometimes f is a very complicated function, or it could be a function of the expected value!
- propagation of errors often not possible \rightarrow resampling schemes
- let $\langle f \rangle$ denote Monte Carlo estimate of some quantity f using all X_k for $k = 1, 2, \dots, N$
- let $\langle f \rangle_J$ denote Monte Carlo estimate of f **omitting** X_J (so use the other $N - 1$ values X_k)

- **jackknife** error estimate given by

$$\sigma^{(J)} = \left(\frac{N-1}{N} \sum_{J=1}^N (\langle f \rangle_J - \langle f \rangle)^2 \right)^{1/2}$$

- Monte Carlo error formula can be used to determine covariance matrix σ_{ij} for correlation function itself in χ^2
- jackknife gives errors in model fit parameters

Bootstrap resampling

- another resampling scheme is the **bootstrap**
- again, let $\langle f \rangle$ denote Monte Carlo estimate of some quantity f using all X_k for $k = 1, 2, \dots, N$
- let $\langle f \rangle_b$ denote Monte Carlo estimate of f using a new set \widehat{X}_k for $k = 1, 2, \dots, N$ where each \widehat{X}_k is one of the original X_j chosen randomly with equal probability (a bootstrap sample)
- a given X_j can occur multiple times in the bootstrap sample
- obtain large number B of such estimates
- let $\widehat{\langle f \rangle} = (1/B) \sum_{b=1}^B \langle f \rangle_b$
- bootstrap error given by

$$\sigma^{(B)} = \left(\frac{1}{B-1} \sum_{b=1}^B (\langle f \rangle_b - \widehat{\langle f \rangle})^2 \right)^{1/2}$$

- plot of probability distribution from bootstrap estimates

The effective mass

- particularly good visual tool to see how well energy extracted is so-called **effective mass**
- for correlator $C(t)$, effective mass defined by

$$m_{\text{eff}}(t) = \ln \left(\frac{C(t)}{C(t + a_t)} \right)$$

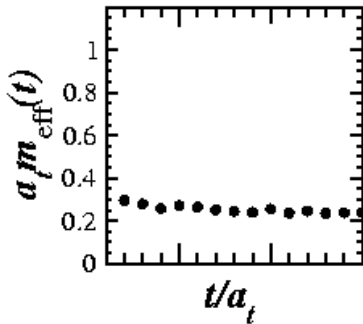
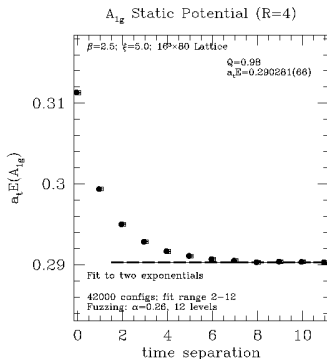
- function which tends to $E_1 - E_0$ as t becomes large

$$\begin{aligned} \lim_{t \rightarrow \infty} m_{\text{eff}}(t) &= \lim_{t \rightarrow \infty} \ln \left(\frac{A_1 e^{-(E_1 - E_0)t} \left(1 + (A_2/A_1) e^{-(E_2 - E_1)t} + \dots \right)}{A_1 e^{-(E_1 - E_0)(t+a_t)} \left(1 + (A_2/A_1) e^{-(E_2 - E_1)(t+a_t)} + \dots \right)} \right) \\ &= \ln \left(e^{(E_1 - E_0)a_t} \right) = a_t (E_1 - E_0). \end{aligned}$$

- value $E_1 - E_0$ seen as large-time **plateau** in effective mass
- contributions from faster-decaying exponentials seen as deviations of the effective mass from its asymptotic plateau value
- “good” operator with little coupling to higher-lying states = rapid onset of plateau
- statistically noise generally grows with t

The effective mass (continued)

- two examples of effective masses
- left: static quark-antiquark potential for separation 0.5 fm
- right: nucleon



Excited states from correlation matrices

- extracting more than just the lowest energy in a symmetry channel requires a hermitian **matrix** of correlation functions $C_{ij}(t)$
- let $\lambda_n(t, t_0)$ denote eigenvalues of $C(t_0)^{-1/2} C(t) C(t_0)^{-1/2}$, for t_0 some fixed reference time
- these eigenvalues can be viewed as **principal** correlators
- ordered such that $\lambda_0 \geq \lambda_1 \geq \dots$ as t becomes large
- can show that

$$\lim_{t \rightarrow \infty} \lambda_n(t, t_0) = e^{-E_n(t-t_0)} \left(1 + O(e^{-\Delta_n(t-t_0)}) \right),$$
$$\Delta_n = \min_{k \neq n} |E_k - E_n|.$$

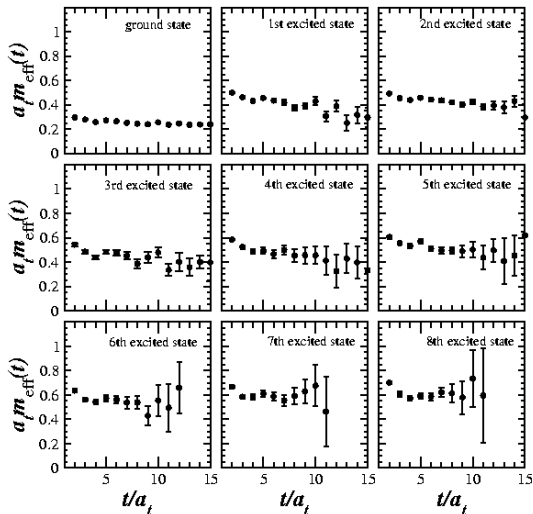
- **principal effective masses** associated with principal correlators

$$m_{\text{eff}}^{(n)}(t) = \ln \left(\frac{\lambda_n(t, t_0)}{\lambda_n(t + a_t, t_0)} \right)$$

- for $N \times N$ correlation matrix, these functions plateau to N lowest lying energies

Principal effective masses

- LHPC currently holds world record for most energy levels extracted in any lattice QCD computation: 9 in nucleon channel



Spectrum for free scalar field theory

- for free-field case on $N_x \times N_y \times N_t$ lattice, define

$$\Phi(t, n_x, n_y) = \sum_{x,y} \phi(x, y, t) e^{2\pi i x n_x / N_x + 2\pi i y n_y / N_y}$$

- lowest six levels having total zero momentum can be extracted using the following set of six operators:

$$O_0(t) = \Phi(t, 0, 0)$$

$$O_1(t) = \Phi(t, 0, 0) \Phi(t, 0, 0)$$

$$O_2(t) = \Phi(t, 1, 0) \Phi(t, -1, 0)$$

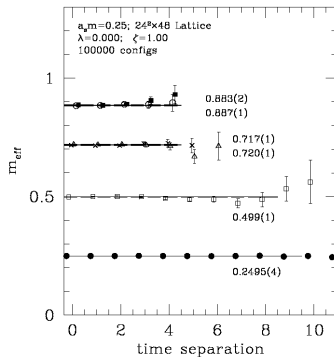
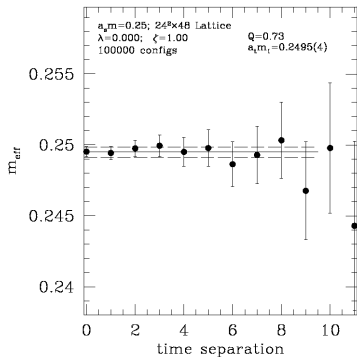
$$O_3(t) = \Phi(t, 0, 1) \Phi(t, 0, -1)$$

$$O_4(t) = \Phi(t, 1, 1) \Phi(t, -1, -1)$$

$$O_5(t) = \Phi(t, 1, -1) \Phi(t, -1, 1)$$

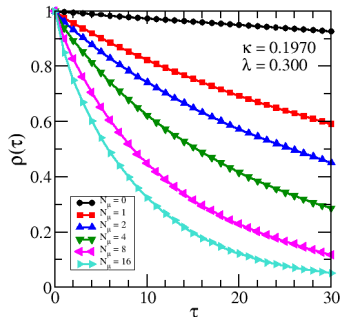
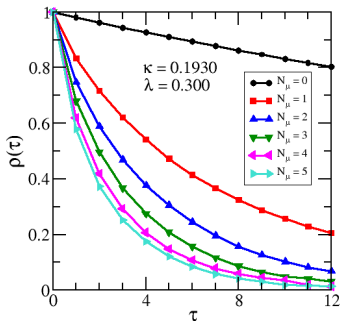
Spectrum for $\lambda = 0$

- extracted six lowest-lying levels in $\lambda = 0$ scalar field theory
- $24^2 \times 48$ isotropic lattice with $a_s m = 0.25$
- exact results: 0.24935 for the mass, 0.49871 for twice the mass, 0.71903 for the two states having minimal relative momenta, and 0.88451 for the next two states



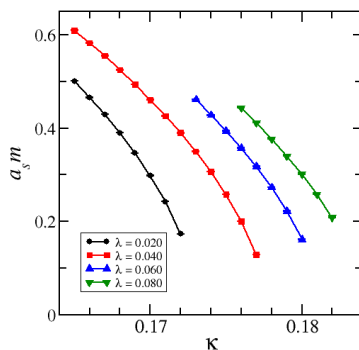
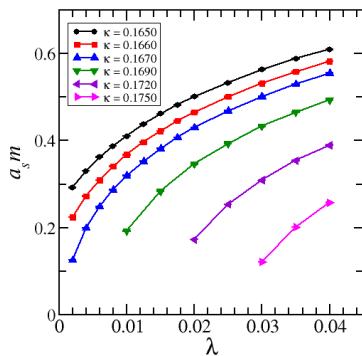
Autocorrelations in the interacting theory

- autocorrelations $\rho(\tau)$ of $\langle \Phi(t)\Phi(0) \rangle$ for $t \sim 1/(2a_s m_{\text{gap}})$
- compound sweep = 1 Metropolis + N_μ microcanonical sweep
- $\mu = 1$ is probability of proposing change in microcanonical
- left plot: $t = 2a_t$ used with $\kappa = 0.1930$ and $\lambda = 0.300$ on $24^2 \times 48$ isotropic lattice and $a_s m_{\text{gap}} \sim 0.25$
- right plot: $t = 5a_t$ used with $\kappa = 0.1970$ and $\lambda = 0.300$ on $32^2 \times 96$ isotropic lattice and $a_s m_{\text{gap}} \sim 0.10$
- microcanonical acceptance rate about 80% in both cases



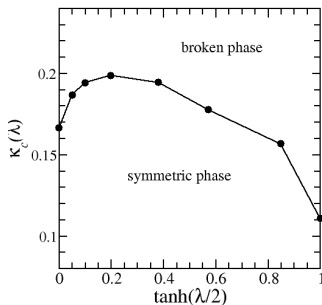
Mass gaps

- various single particle masses on 24^3 isotropic lattice



Phase structure

- theory has two phases separated by a line of critical points
- for each value of λ , there exists a critical value $\kappa_c(\lambda)$ at which mass gap goes to zero
- symmetric phase for $\kappa < \kappa_c(\lambda)$
 - $\phi \rightarrow -\phi$ symmetry holds, $\langle \phi \rangle = 0$
- broken phase for $\kappa > \kappa_c(\lambda)$
 - $\phi \rightarrow -\phi$ spontaneously broken, $\langle \phi \rangle \neq 0$

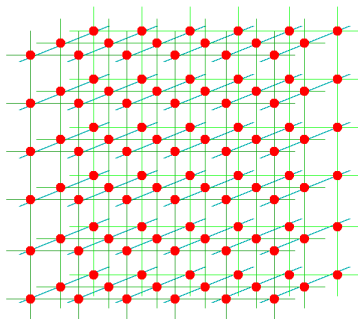


Part V

Monte Carlo calculations in lattice Quantum Chromodynamics

Lattice QCD

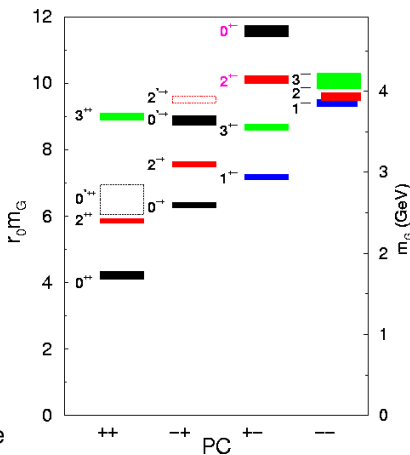
- hypercubic space-time lattice
- **quarks** reside on sites, **gluons** reside on links between sites
- for gluons, 8 dimensional integral on each link
- path integral has dimension $32N_x N_y N_z N_t$
 - 10.6 million for 24^4 lattice
- more sophisticated updating algorithms
- systematic errors
 - discretization
 - finite volume



Glueball spectrum in pure gauge theory

- gluons can bind to form **glueballs**
 - e.m. analogue: massive globules of pure light!
- states labeled by J^{PC}
- scale set by $r_0^{-1} = 410(20)$ MeV
- computed using pseudo-heatbath and microcanonical
- 24×24 correlation matrix in each symmetry channel
- spin identification
- mass gap with a bounty
 - Clay mathematics institute will pay \$ 1 million

C. Morningstar and M. Peardon,
Phys. Rev. D 60, 034509 (1999)



Conclusion

- observables in quantum mechanical systems can be extracted from the correlation functions of the theory
- correlation functions can be computed using path integrals
- path integrals in the imaginary time formalism can be evaluated using the Monte Carlo method
- importance sampling from Markov chains
- Metropolis-Hastings method
- microcanonical updating
- 1-dimensional simple harmonic oscillator was first example
- calculations in real scalar ϕ^4 theory in $2 + 1$ dimensions

For Further Reading

-  C.M. Grinstead and J.L. Snell,
Introduction to Probability
-  E. Parzen, *Stochastic Processes*
(Holden-Day, San Francisco, 1962).
-  N.U. Prabhu, *Stochastic Processes*
(Macmillan, New York, 1965).
-  I. Montvay and G. Münster
Quantum Fields on a Lattice
(Cambridge Press, 1994).
-  S. Karlin and H. Taylor,
A First Course in Stochastic Processes
(Academic Press, 1975).
-  J. Hamilton, *Time Series Analysis*
(Princeton University Press, 1994).